

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

#### 4 protein - protein search, using sw model

run on: June 8, 2004, 13:55:43 ; Search time 91.7861 Seconds  
(without alignments)  
1065.101 Million cell updates/sec

title: US-10-010-050A-2

sequence: 1 MRGAGARGASWCMALAL.....IKITYEIPLPFRMKTSGL 346

scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

post-processing: Minimum March 0%  
Maximum March 100%  
Listing first 45 summaries

database: A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1920	99.6	346	2	AAW92967 Human zsi
2	1916	99.4	347	2	AAV41323 Human sec
3	1913	99.3	358	4	AAW93870 Human pol
4	886	46.0	163	4	AAW19604 Peptide #
5	886	46.0	163	4	ABH39239 Peptide #
6	886	46.0	163	4	AAW32740 Peptide #
7	886	46.0	163	4	ABH24080 Protein #
8	886	46.0	163	4	AAW72492 Human bon
9	886	46.0	163	4	AAW59901 Human bra
10	886	46.0	163	4	ABG54183 Human liv
11	886	46.0	163	5	ABG42312 Human pep
12	472	24.5	89	5	ABP42902 Human ova
13	96.5	5.0	1705	5	ABR52348 Protein r
14	95.5	5.0	2703	4	ABH60074 Drosophil
15	94	4.9	429	5	ABU52308 Helicobac
16	94	4.9	1238	6	ABU30825 Protein e
17	93	4.8	478	6	ABU27691 Protein e
18	92.5	4.8	569	4	ABH71458 Drosophil
19	91	4.7	433	3	AAW31636 Arabidops
20	91	4.7	440	2	AAW41852 Ragweed p
21	91	4.7	508	3	AAW31635 Arabidops
22	91	4.7	742	3	AAW94977 Human sec
23	91	4.7	742	7	ADH56938 Human pro
24	91	4.7	3923	2	AAW31237 Human Apo
25	91	4.7	4563	6	ABU79140 Apolipop

26	91	4.7	4590	4	AAU33184 Novel hum
27	89.5	4.6	442	2	AAW41856 Ragweed p
28	89.5	4.6	2237	5	ABG70004 Larval vi
29	89	4.6	568	4	ABH60986 Drosophil
30	88.5	4.6	1462	7	ADD14088 Human src
31	88.5	4.6	1468	4	AAW39218 Human pol
32	88.5	4.6	1496	4	AAW41004 Human pol
33	88	4.6	605	6	ABU32174 Protein e
34	88	4.6	3194	6	ABU31122 Protein e
35	87.5	4.5	311	4	AAH06730 Human CAS
36	87	4.5	226	3	ABH34827 Gene 17 h
37	87	4.5	502	2	AAW58861 T. haloph
38	87	4.5	661	2	AAW15565 Human int
39	86.5	4.5	388	5	ABH92540 Hericida
40	86	4.5	411	5	ABH49050 listeria
41	86	4.5	411	6	ABU32653 Protein e
42	85.5	4.4	428	5	AAU74922 Synthetic
43	85.5	4.4	5035	2	AAW25450 MH mutant
44	84.5	4.4	1462	2	AAW37508 Human DNA
45	83.5	4.3	373	3	AAH52581 Helicobac

#### ALIGNMENTS

RESULT 1  
ID AAW92967 standard; protein: 346 AA.  
XX  
AC AAW92967;  
XX  
DT 14-MAY-1999 (first entry)  
XX  
DE Human zsig46 protein.  
XX  
KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;  
KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;  
KW Hirschsprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;  
KW Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;  
KW therapy; diagnostic.  
XX  
OS Homo sapiens.  
XX  
PN W09905275-A1.  
XX  
PD 04-FEB-1999.  
XX  
PF 24-JUL-1998; 98WO-US015431.  
XX  
PR 24-JUL-1997; 97US-0053613P.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Sheppard PO, Gilbertson DG;  
XX  
DR WPI; 1999-142930/12.  
XX  
DR N-PSDB; AAX02855.  
XX  
PT New secreted polypeptide, zsig46, and its fragments, related fusion  
PT proteins - used for diagnosis and treatment of thyroid disorders or  
PT diseases involving genes on chromosome 13.  
XX  
PS Claim 3; Page 90-91; 101pp; English.  
XX  
CC This invention describes the isolation of a novel human secreted protein,  
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in  
CC the thyroid. This product can be used to study secretion of proteins from  
CC cells and also to treat or prevent deficient expression of zsig46, which  
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
CC disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger  
CC syndrome). Antibodies and other binding proteins, are used as immunoassay

```

1C reagents to detect zsi46 or cells expressing it, e.g. for assessing
1C thyroid function to produce anti-idiotypic antibodies, for affinity
1C purification of zsi46, to screen expression libraries, to neutralise
1C zsi46 activity, and to deliver toxins, radioscopes etc. for
1C therapeutic or diagnostic purposes. Agonists of the product can be used
1C to promote growth, differentiation and proliferation of specific cell
1C types, e.g. for treating (extra)thyroid diseases or as additive to cell
1C cultures
1C
1C Sequence 346 AA;
1C
1C Query Match          99.6%; Score 1920; DB 2; Length 346;
1C Best Local Similarity 99.7%; Pred. No. 4.2e-185;
1C Matches 345; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

2y	1	MERGGAARGRASWOMALALMLAVPMSVSSGIPERRHMPVYKFDPRPDDYCA	60
2b	1	MORGGAARGRASWOMALALMLAVPMSVSGIPERRHMPVYKFDPRPDDYCA	60
2y	61	KTTECPPTSGPIPVMEGGDDIVFRLQVAFWEFKYGLGHKIMHDAIGRSTLTGKNT	120
2b	61	KYTPCPPTSGPIPVMEGGDDIVFRLQVAFWEFKYGLDLGHKIMHDAIGRSTLTGKNT	120
2y	121	MEWYELFOLGNTCFPHLRPMDADPWCOCAGACEFEGGIDVHWKENGTLVQVATISGNE	160
2b	121	MEWYELFOLGNTCFPHLRPMDADPWCOCAGACEFEGIDVHWKENGTLVQVATISGNE	160
2y	181	NOMAKWVQDNDETGIYYETANVAKASPEKGATWDSYDSCSFVLRTNNKLABGAERKNT	240
2b	181	NOMAKWVQDNDETGIYYETANVAKASPEKGATWDSYDSCSFVLRTNNKLABGAERKNT	240
2y	241	ETNYRIRIFLYSEGPYLLGNETSVPQPTNNKTLGLAIRFYYPFKPHLPTKEFLSLILOIF	300
2b	241	ETNYRIRIFLYSEGPYLLGNETSVPQPTNNKTLGLAIRFYYPFKPHLPTKEFLSLILOIF	300
2y	301	DAVIYHKOPLYLYNFYEWFLPMKRPFKITTEELPLPRNNKTLISGL	346
2b	301	DAVIYHKOPLYLYNFYEWFLPMKRPFKITTEELPLPRNNKTLISGL	346

RESULT 2	
AAV41323	
ID	AAV41323 standard; protein; 347 AA.
XX	
AC	AAV41323;
XX	
DT	02-DEC-1999 (first entry)
XX	
	Human secreted protein encoded by gene 16 clone HWZAD77

Human: secreted protein; fusion protein; gene therapy; protein therapy;  
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;  
developmental abnormality; foetal deficiency; blood; allergy; renal;  
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
inflammation; ischaemic shock; Alzheimer's disease; osteosarcoma; AIDS;  
cognitive disorder; schizophrenia; prostatic disease; osteoclast; thymus;  
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm;

OS	io:omo sapiens.
XX	
PN	MO947540-AL.
PD	
XX	23-SEP-1999.
PF	
XX	18-MAR-1999;
XX	99MO-US005804.
PR	19-MAR-1999;
PR	98US-0078563P
PR	19-MAR-1998;
PR	98US-0078566P
PR	18-MAR-1998;
PR	98US-0078573P
PR	19-MAR-1998;
PR	98US-0078574P
PR	19-MAR-1998;
PR	98US-0078575P
PR	19-MAR-1998;
PR	98US-0078576P
PR	19-MAR-1998;
PR	98US-0078577P
PR	19-MAR-1998;
PR	98US-0078578P
PR	19-MAR-1998;
PR	98US-0078579P
PR	19-MAR-1998;
PR	98US-0078580P
PR	19-MAR-1998;
PR	98US-0078581P
PR	19-MAR-1998;
PR	98US-0078582P
PR	19-MAR-1998;
PR	98US-0078583P
PR	19-MAR-1998;
PR	98US-0078584P
PR	19-MAR-1998;
PR	98US-0078585P
PR	19-MAR-1998;
PR	98US-0078586P
PR	19-MAR-1998;
PR	98US-0078587P
PR	19-MAR-1998;
PR	98US-0078588P
PR	19-MAR-1998;
PR	98US-0078589P
PR	19-MAR-1998;
PR	98US-0078590P
PR	19-MAR-1998;
PR	98US-0078591P
PR	19-MAR-1998;
PR	98US-0078592P
PR	19-MAR-1998;
PR	98US-0078593P
PR	19-MAR-1998;
PR	98US-0078594P
PR	19-MAR-1998;
PR	98US-0078595P
PR	19-MAR-1998;
PR	98US-0078596P
PR	19-MAR-1998;
PR	98US-0078597P
PR	19-MAR-1998;
PR	98US-0078598P
PR	19-MAR-1998;
PR	98US-0078599P
PR	19-MAR-1998;
PR	98US-0078600P
PR	19-MAR-1998;
PR	98US-0078601P
PR	19-MAR-1998;
PR	98US-0078602P
PR	19-MAR-1998;
PR	98US-0078603P
PR	19-MAR-1998;
PR	98US-0078604P
PR	19-MAR-1998;
PR	98US-0078605P
PR	19-MAR-1998;
PR	98US-0078606P
PR	19-MAR-1998;
PR	98US-0078607P
PR	19-MAR-1998;
PR	98US-0078608P
PR	19-MAR-1998;
PR	98US-0078609P
PR	19-MAR-1998;
PR	98US-0078610P
PR	19-MAR-1998;
PR	98US-0078611P
PR	19-MAR-1998;
PR	98US-0078612P
PR	19-MAR-1998;
PR	98US-0078613P
PR	19-MAR-1998;
PR	98US-0078614P
PR	19-MAR-1998;
PR	98US-0078615P
PR	19-MAR-1998;
PR	98US-0078616P
PR	19-MAR-1998;
PR	98US-0078617P
PR	19-MAR-1998;
PR	98US-0078618P
PR	19-MAR-1998;
PR	98US-0078619P
PR	19-MAR-1998;
PR	98US-0078620P
PR	19-MAR-1998;
PR	98US-0078621P
PR	19-MAR-1998;
PR	98US-0078622P
PR	19-MAR-1998;
PR	98US-0078623P
PR	19-MAR-1998;
PR	98US-0078624P
PR	19-MAR-1998;
PR	98US-0078625P
PR	19-MAR-1998;
PR	98US-0078626P
PR	19-MAR-1998;
PR	98US-0078627P
PR	19-MAR-1998;
PR	98US-0078628P
PR	19-MAR-1998;
PR	98US-0078629P
PR	19-MAR-1998;
PR	98US-0078630P
PR	19-MAR-1998;
PR	98US-0078631P
PR	19-MAR-1998;
PR	98US-0078632P
PR	19-MAR-1998;
PR	98US-0078633P
PR	19-MAR-1998;
PR	98US-0078634P
PR	19-MAR-1998;
PR	98US-0078635P
PR	19-MAR-1998;
PR	98US-0078636P
PR	19-MAR-1998;
PR	98US-0078637P
PR	19-MAR-1998;
PR	98US-0078638P
PR	19-MAR-1998;
PR	98US-0078639P
PR	19-MAR-1998;
PR	98US-0078640P
PR	19-MAR-1998;
PR	98US-0078641P
PR	19-MAR-1998;
PR	98US-0078642P
PR	19-MAR-1998;
PR	98US-0078643P
PR	19-MAR-1998;
PR	98US-0078644P
PR	19-MAR-1998;
PR	98US-0078645P
PR	19-MAR-1998;
PR	98US-0078646P
PR	19-MAR-1998;
PR	98US-0078647P
PR	

PR 19-MAR-1998; 98US-0078578P.  
PR 19-MAR-1998; 98US-0078579P.  
PR 19-MAR-1998; 98US-0078581P.  
PR 01-APR-1998; 98US-0080312P.  
PR 01-APR-1998; 98US-0080313P.  
PR 01-APR-1998; 98US-0080314P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR; ...

PI Shi Y, Moore PA;

DR WPI; 1999-562050/47.

XX N-EDD, XXXX-XXXX.

XX

PT cancers, neurological disorders, immune diseases, inflammation or blood

PT disorders.  
XX

PS Claim 11; page 36/-368; 484pp; English.  
XX

CC This sequence represents a secreted human protein encoded by the gene  
CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin HC portion (e.g. HAZ4802) for increasing the stability of the fused protein.

CC as compared to the human protein only. The invention relates to 95 novel  
CC clones and their fragments (nucleic acid sequences: AA224811-724907; ami-

CC acid sequences AAY41308-Y41404) which are useful for preventing, treating, or diagnosing a disease or condition, or for protein or gene therapy.

CC pathological conditions can be diagnosed by determining the amount of the

CC in the new polynucleotides. Specific uses are described for each of the

CC in (see AA224811 for described uses)

SQ Sequence 347 AA;

Query Match 99.48; Score 1916; DB 2; Length 347;

```

      0; Mismatches      2; Indels      0; Gaps      0;
Matches 344; Conservative

```

1 MRGAGARGRASWCALALTLAVVPGMSRVSGIPSRHWPVPYKRFDPRKPDYCA 60

Db 1 MRGAGAGRGASWCWALALMLAVVPGMSRVSGIPSPRHMPVPEYKRFDERPKPDYCOA 60

61 KYTECPGSPVMEGDDIEVERLOAPVWEFKYGDLLGHLKIMHDAIGERSTLTGKNYT 12

61 KYTECP TGSPI PYMEGDDDI EYERIQAPVWEFKYGLTGLKIMHDAIGERSTLTGKNT 122

1 21 MEWYET.FOI.GNCTEDHI.PBEMDABFWCNOGAA.CFEFEGT.DVHWKENG.TI.VOVATISGNME 18

131 MEWYEI FOI GNCTEBYI BREMDA BEWCNOGA ACEEFGIDYHWKENGTI VOYATISGNME 18

101 NOMA KATYKODNIETCTVETEMNIIKASPEKCAETWENDCSKEYI PTENKI AEEGAEEKNT 24

[illegible]

THE UNIVERSITY OF CHICAGO PRESS

SECRET

1000

[illegible][illegible]

### RESULT 3

AAM93870  
ID AAM93870 standard; protein; 358 AA.

AA  
AC  
AAM93870;



```

06-NOV-2001 (first entry)
Human polypeptide, SEQ ID NO: 3978.
Human; full length cDNA; cDNA synthesis; oligo-capping.
Homo sapiens.
EP130094-A2.
05-SEP-2001.
07-JUL-2000; 2000EP-00114089.
08-JUL-1999; 99JP-00194486.
11-JAN-2000; 2000JP-00118774.
02-MAY-2000; 2000JP-00183765.
(HELI-) HELIX RES INST.
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
WPI; 2001-524255/58.
N-PSDB; AAK94829.
830 Primers useful for synthesizing full length cDNA clones and their use
in genetic manipulation.
Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.
The invention relates to primers for synthesizing full length cDNA
clones. 830 cDNA molecules encoding a human protein have been isolated
and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
been determined. Primers for synthesizing the full length cDNA are useful
for clarifying the function of the protein encoded by the cDNA. The full
length clones were obtained by construction of full length enriched cDNA
libraries that were synthesised by the oligo-capping method. The primers
enable the production of the full length cDNA easily without any special
methods. The present sequence is a polypeptide encoded by a full length
human cDNA of the invention. Note: The sequence data for this patent did
not form part of the printed specification, but was obtained in CD-ROM
format directly from EPO
Sequence 358 AA;
Query Match 99.3%; Score 1913; DB 4; Length 358;
Best Local Similarity 99.4%; Pred. No. 2.2e-184;
Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 MRGGAARAGRAAGCAGLALMLAVVPGMSKVSIGIPSRHHVYPRKRPFRKPPYCOA 60
13 MRGGAARAGRAAGCAGLALMLAVVPGMSKVSIGIPSRHHVYPRKRPFRKPPYCOA 72
61 KTFPGTSGPIVMEGDDDDIEVFRLOAPYMEFKYDGLGHLKIMDAIGFRSTLGGKNT 120
73 KTFPGTSGPIVMEGDDDDIEVFRLOAPYMEFKYDGLGHLKIMDAIGFRSTLGGKNT 132
121 MEMYELFOLGNCITPHLRREMDAPFMCNGAACFFEGIDDVHMKNGTLVQVATISGNMF 180
133 MEMYELFOLGNCITPHLRREMDAPFMCNGAACFFEGIDDVHMKNGTLVQVATISGNMF 192
181 NOMAAWVQNDNTGTYETWNVKASPEKGAETWPSYDSCSKVATFNKLAFFGAEPKRI 240
193 NOMAAWVQNDNTGTYETWNVKASPEKGAETWPSYDSCSKVATFNKLAFFGAEPKRI 252
241 EFNVYRIFLYSEPTLYLGNETSVPFGPTGKTLGLAIKRFYYPFKPHLPKXEFLLSLQIF 300
253 EFNVYRIFLYSEPTLYLGNETSVPFGPTGKTLGLAIKRFYYPFKPHLPKXEFLLSLQIF 312
301 DAVIVKQFYLFYNEFWFLPMKPPFIKITTYEELPIPIRNKTLISGL 346
313 DAVIVKQFYLFYNEFWFLPMKPPFIKITTYEELPIPIRNKTLISGL 358

```

```

RESULT 4
ID AA019604
AA019604 standard; protein, 163 AA.
AC AA019604;
XX 12-OCT-2001 (first entry)
XX DE Peptide #6038 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX OS Homo sapiens.
XX PN MO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632386.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX PS Claim 27; SEQ ID NO 24430; 487bp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX (SENP: see AA110068-AT28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 163 AA;
Query Match 46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
184 AKWVQNDNETGTYETWNVKASPEKGAETWPSYDSCSKVATFNKLAFFGAEPKRI 243
1 AKWVQNDNETGTYETWNVKASPEKGAETWPSYDSCSKVATFNKLAFFGAEPKRI 60
244 YTRIFLYSGEPTLYLGNETSVPFGPTGKTLGLAIKRFYYPFKPHLPKXEFLLSLQIF 303
61 YTRIFLYSGEPTLYLGNETSVPFGPTGKTLGLAIKRFYYPFKPHLPKXEFLLSLQIF 312
304 IVHKOYFLFYNEFWFLPMKPPFIKITTYEELPIPIRNKTLISGL 346
121 IVHKOYFLFYNEFWFLPMKPPFIKITTYEELPIPIRNKTLISGL 163
RESULT 5

```

```

ABB39239
ID ABB39239 standard; peptide; 163 AA.
CX ABB39239;
YC
YT 04-FEB-2002 (first entry)
ZE Peptide #6745 encoded by human foetal liver single exon probe.
ZW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
YN WC00157277-A2.
XD 09-AUG-2001.
XE 30-JAN-2001; 2001MO-US000669.
XF 04-FEB-2000; 2000US-0180312P.
XG 26-MAY-2000; 2000US-0207456P.
XH 30-JUN-2000; 2000US-00608408.
XI 03-APR-2000; 2000US-00632366.
XJ 21-SEP-2000; 2000US-0234687P.
XK 27-SEP-2000; 2000US-0236359P.
XL 04-OCT-2000; 2000GB-00024263.
XM (MOLE-) MOLECULAR DYNAMICS INC.
XN Penn SG, Hanzel DX, Chen W, Rank DR;
YO WPI; 2001-483447/52.
YX
YZ Human genome-derived single exon nucleic acid probes useful for analyzing
ZA gene expression in human fetal liver.
ZB Claim 27; SEQ ID NO 31874; 639bp + Sequence Listing; English.
ZC The invention relates to a single exon nucleic acid probe for measuring
ZE human gene expression in a sample derived from human foetal liver. The
ZF single exon nucleic acid probes may be used for predicting, measuring and
ZH displaying gene expression in samples derived from human fetal liver. The
ZI present sequence is a peptide encoded by a single exon nucleic acid probe
XJ of the invention. Note: The sequence data for this patent did not form
XC part of the printed specification, but was obtained in electronic format
XD directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XE
XF Sequence 163 AA;
XG
XH Query Match 46.0%; Score 886; DB 4; Length 163;
XH Best Local Similarity 100.0%; Pred. No. 4.ee-01;
XH Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0
YI
YJ 184 AKWVKQDNETGIYYETNVVAKASPEKGATWDSDSIDCSKFVIRTNKLAERGAEFKNIEFN 243
YK 1 AKWVKQDNETGIYYETNVVAKASPEKGATWDSDSKFVIRTNKLAERGAEFKNIEFN 60
YL 244 YRRIFLYSGEPFYLLGNETSVEGPNTNKTLGAIKRFFYPFPKHPTKEFTLSLLOTFDAY 303
YM 61 YRRIFLYSGEPFYLLGNETSVEGPNTNKTLGAIKRFFYPFPKHPTKEFTLSLLOTFDAY 120
YN 304 IVHKQFLPFYNFEYWFLPMKKPFPIKITYEELIPDIRNKTLISGL 346
YO 121 IVHKQFLPFYNFEYWFLPMKKPFPIKITYEELIPDIRNKTLISGL 163
YX
YD RESULT 6
YD AAM32740
YD AAM32740 standard; protein; 163 AA.
YC
YX AAM32740;
YT 17-OCT-2001 (first entry)

```

XX		Peptide #6777 encoded by probe for measuring placental gene expression.
XX		
XX	KM	Probe; microarray; human; placenta; antenatal diagnosis;
XX	KM	genetic disorder.
OS	Homo sapiens.	
XX		
XX	PN	MO200157272-A2.
PD		
XX		09-AUG-2001.
PF		
XX		30-JAN-2001; 2001WO-US0000663.
XX		
PR		04-FEB-2000; 2000US-0180312P.
XX		
PR		26-MAY-2000; 2000US-0207456P.
XX		
PR		30-JUN-2000; 2000US-00608408.
XX		
PR		03-AUG-2000; 2000US-00623256.
XX		
PR		21-SEP-2000; 2000US-0234687P.
XX		
PR		27-SEP-2000; 2000US-0236359P.
XX		
PR		04-OCT-2000; 2000GB-00024263.
PA	(MOLE-) MOLECULAR DYNAMICS INC.	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-488897/53.	
XX		
PT	Human genome-derived single exon nucleic acid probes useful for analyzing	
XX	gene expression in human placenta.	
PS	Claim 27; SEQ ID NO 33009; 654bp; English.	
CC	The present invention relates to single exon nucleic acid probes (SNP:	
XX	see AAI33115-AAI57546). The present sequence is a peptide encoded by one	
CC	such probe. The probes are useful for producing a microarray for	
XX	predicting, measuring and displaying gene expression in samples derived	
CC	from human placenta. The probes are useful for antenatal diagnosis of	
CC	human genetic disorders	
SQ	Sequence 163 AA:	
	Query Match	46.0%; Score 886; DB 4; Length 163;
	Best local Similarity	100.0%; Pred. No. 4.8e-81;
	Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	184 AKWKYKODNETHSYVETNNVKKSPKGAETWPDSYDCKPVLTFFPKLAFFGAEFKNIEN	243
DB	1 AKWKYKODNETGIYYETNNVKASPEKGATWPDFSDYSKFVLTFPKLAEFGAEFKNIETN	60
OY	244 YTRIFLVSGEFTYIGNETSVEFPTGNKTGLAIRFYYPFEKHLPTKEELSLIQITFDAY	303
DB	61 YTRIFLVSGEFTYIGNETSVEFPTGNKTGLAIRFYYPFEKHLPTKEELSLIQITFDAY	120
OY	304 IVHKQFTLFYNFEYFWFLPMKPPFIKITTEBILPLIRNKTLSEL	346
DB	121 IVHKQFTLFYNFEYFWFLPMKPPFIKITTEBILPLIRNKTLSEL	163
RESULT 7		
ABB24080		
ID	ABB24080 standard; protein; 163 AA.	
AC	ABB24080;	
XX		
DT	23-JAN-2002 (first entry)	
XX		
DE	Protein #6079 encoded by probe for measuring heart cell gene expression.	
XX		
KM	Human; gene expression; heart; microarray; vascular system;	
KM	cardiovascular disease; hypertension; cardiac arrhythmia;	
KM	congenital heart disease.	
XX		

```

3 Homo sapiens.
4 WO200157276-A2.
5
6 09-AUG-2001.
7
8 30-JAN-2001; 2001WO-US000666.
9
10 04-FEB-2000; 2000US-0180312P.
11 26-MAY-2000; 2000US-0207456P.
12 30-JUN-2000; 2000US-00608408.
13 03-AUG-2000; 2000US-00632366.
14 21-SEP-2000; 2000US-0234687P.
15 27-SEP-2000; 2000US-0236359P.
16 04-OCT-2000; 2000GB-00024263.
17
18 (MOLE-) MOLECULAR DYNAMICS INC.
19 Penn SG, Hanzel DK, Chen W, Rank DR;
20 WPI; 2001-488900/53.
21
22 Single exon nucleic acid probes for analyzing gene expression in human
23 hearts.
24
25 Claim 15; SEQ ID NO 25850; 530bp; English.
26
27 The present invention relates to single exon nucleic acid probes for
28 measuring human gene expression in a sample derived from human heart (see
29 ABA21533-AM41305). The present sequence is a protein encoded by one such
30 probe. The probes may be used for predicting, measuring and displaying
31 gene expression in samples derived from the human heart via microarrays.
32 By measuring gene expression, the probes are useful for predicting,
33 diagnosing, grading, staging, monitoring and prognosing diseases of the
34 human heart and vascular system e.g. cardiovascular disease,
35 hypertension, cardiac arrhythmias and congenital heart disease. Note: The
36 sequence data for this patent did not form part of the printed
37 specification, but was obtained in electronic format directly from WIPO
38 at ftp.wipo.int/pub/published_pct_sequences
39
40 Sequence 163 AA:
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
```

```

R 04-FEB-2000; 2000US-0180312P.
R 26-MAY-2000; 2000US-0207456P.
R 30-JUN-2000; 2000US-00608408.
R 03-AUG-2000; 2000US-00632366.
R 21-SEP-2000; 2000US-0234687P.
R 27-SEP-2000; 2000US-0236359P.
R 04-OCT-2000; 2000GB-00024263.
X A (MOLE-) MOLECULAR DYNAMICS INC.
X I Penn SG, Hanzel DK, Chen W, Rank DR;
X R WPI; 2001-483446/52.
X T Single exon nucleic acid probes for analyzing gene expression in human
X T brains.
X S Example 4; SEQ ID NO 32006; 650bp + Sequence Listing; English.
X S The present invention provides a number of single exon nucleic acid
X C probes which are derived from genomic sequences expressed in the human
X C brain. They can be used to measure gene expression in brain cell samples,
X C which may enable the diagnosis and improved treatment of nervous system
X C diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
X C epilepsy and cancers. The present sequence is a protein encoded by one of
X C the probes of the invention
X Q Sequence 163 AA:

Query Match          46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 184 AKWVKODNETGYIETWNNVYASPEKGAETWFDSDSKVLRFTFNKLAERGAFFKNIETN 243
b 1 AKWVKODNETGYIETWNNVYASPEKGAETWFDSDSKVLRFTFNKLAERGAFFKNIETN 60
Y 244 YTRIFLYSGEPYLLGNETSVEFGPTGNKTLGLAIKRFYFFPKPLPTKEFLLSLQIFDAV 303
b 61 YTRIFLYSGEPYLLGNETSVEFGPTGNKTLGLAIKRFYFFPKPLPTKEFLLSLQIFDAV 120
Y 304 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIPLPINKNTLSGL 346
b 121 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIPLPINKNTLSGL 163

ESULT 10
BSG54183
D ABG54183 standard; peptide; 163 AA.
C ABG54183;
T 25-FEB-2003 (first entry)
X Human liver peptide, SEQ ID NO 32831.
X Human liver peptide, SEQ ID NO 32831.
X Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
X M hypercholesterolaemia; coronary heart disease.
X S Homo sapiens.
X N WO200157273-A2.
X D 09-AUG-2001.
X F 30-JAN-2001; 2001WO-US000664.
X R 04-FEB-2000; 2000US-0180312P.
X R 26-MAY-2000; 2000US-0207456P.
X R 30-JUN-2000; 2000US-00608408.
X R 03-AUG-2000; 2000US-00632366.
X R 21-SEP-2000; 2000US-0234687P.
X R 27-SEP-2000; 2000US-0236359P.
X R 04-OCT-2000; 2000US-0236359P.

```

```

PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488998/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 32831; 658bp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 1109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABG47348-ABG59930 represent human
XX CC liver single exon encoded peptides of the invention. Note: The sequence
XX CC information for this patent does not appear in the printed specification
XX CC but was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 163 AA:

Query Match          46.0%; Score 886; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.8e-81;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 AKWVKODNETGYIETWNNVYASPEKGAETWFDSDSKVLRFTFNKLAERGAFFKNIETN 243
Db 1 AKWVKODNETGYIETWNNVYASPEKGAETWFDSDSKVLRFTFNKLAERGAFFKNIETN 60
QY 244 YTRIFLYSGEPYLLGNETSVEFGPTGNKTLGLAIKRFYFFPKPLPTKEFLLSLQIFDAV 303
Db 61 YTRIFLYSGEPYLLGNETSVEFGPTGNKTLGLAIKRFYFFPKPLPTKEFLLSLQIFDAV 120
QY 304 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIPLPINKNTLSGL 346
Db 121 IVHKQFYLFYNEFYWFLPMKFPPIKITTYEIPLPINKNTLSGL 163

RESULT 11
ABG42312
ID ABG42312 standard; peptide; 163 AA.
XX ABG42312;
AC 19-AUG-2002 (first entry)
DT Human peptide encoded by genome-derived single exon probe SEQ ID 31977.
XX Human; single exon probe; asthma; lung cancer; COPD; IHD;
XX KM chronic obstructive pulmonary disease; interstitial lung disease;
XX KM familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KM Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;
XX KM pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
XX KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KM primary ciliary dyskinesia; pulmonary hypertension;
XX KM hyaline membrane disease.
XX OS Homo sapiens.
XX OS WO200186003-A2.
XX PN 15-NOV-2001.
XX PD
XX

```

30-JAN-2001; 2001WO-US000665.  
 04-FEB-2000; 2000US-0180312P.  
 26-MAY-2000; 2000US-0207456P.  
 30-JUN-2000; 2000US-00608408.  
 03-AUG-2000; 2000US-00632366.  
 21-SEP-2000; 2000US-0234687P.  
 27-SEP-2000; 2000US-0236359P.  
 04-OCT-2000; 2000GB-00024263.  
 (MOLE-) MOLECULAR DYNAMICS INC.  
 Penn SG, Hanzel DK, Chen W, Rank DR;  
 WPI; 2002-114183/15.  
 Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples.  
 Claim 27; SEQ ID NO 31977; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample delivered from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsageners syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane diseases. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 163 AA;

Query Match 46.0%; Score 886; DB 5; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-81;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

184 AKWKQDNETGTYIYETWNVKASPEKGAETWDSYDCSFVLRITNKLAEFGAEKNTETN 243  
 1 AKWKQDNETGTYIYETWNVKASPEKGAETWDSYDCSFVLRITNKLAEFGAEKNTETN 60  
 244 YTRIFLYSGEPTVYIGNETSVFGPTGNKTLGAIKRFYYPKPHLPYKFLISLIQIFDAV 303  
 61 YTRIFLYSGEPTVYIGNETSVFGPTGNKTLGAIKRFYYPKPHLPYKFLISLIQIFDAV 120

CY 304 IVKQFYLTFNFEYWPFLPMKPFKITTEYELPLPRNKTLSGL 346  
 DB 121 IVKQFYLTFNFEYWPFLPMKPFKITTEYELPLPRNKTLSGL 163

RESULT 12  
 ABP42902  
 ID ABP42902 standard; protein, 69 AA.  
 XX ABP42902;  
 AC  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Human ovarian antigen HPDQ81, SEQ ID NO:4034.

Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
 inflammatory condition; immune disorder; blood disorder;  
 cardiovascular disorder; respiratory disorder; neurological disorder;  
 gastrointestinal disorder; urinary system disorder; drug screening;  
 gene therapy; chromosome mapping; forensic analysis;  
 antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 antiinflammatory; gynaecological; reproductive.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PV WO200200677-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US018569.  
 XX  
 PR 07-JUN-2000; 2000US-0209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-147878/19.  
 DR N-PSDB; ABQ55979.  
 XX

Isolated nucleic acid molecules encoding novel ovarian polypeptides, useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases.  
 PT  
 PT  
 XX

Claim 11; SEQ ID NO 4034; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovarian and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the

identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 89 AA;

Query Match 24.5%; Score 472; DB 5; Length 89;  
Best Local Similarity 98.9%; Pred. No. 1,4e-39;  
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

258 GNETSVGPNTKLTGLAKRFPYPPKPHLPTKEFLLSLQIPDAVYVHKQFLFYNFEX 317  
1 GNETSVGPNTKLTGLAKRFPYPPKPHLPTKEFLLSLQIPDAVYVHKQFLFYNFEX 60

318 WFLPMKEPFIKITYEELPLIRNKTLISGL 346  
61 WFLPMKEPFIKITYEELPLIRNKTLISGL 89

RESULT 13  
ABR52348  
ID ABR52348 standard; protein; 1705 AA.

ABR52348;  
19-JUN-2003 (first entry)

Protein relating to the invention SEQ ID NO: 37.

antiproliferative; hepatotropic; nephrotropic; antiarthritic;  
antiproliferative; cardiac; cytostatic; gene therapy; liver disease;  
proliferative disorder; renal failure; cardiovascular disorder;  
immunological disorder; arthritis; psoriasis; congenital heart defect;  
congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.

Mus musculus.  
WO200257460-A2.  
25-JUL-2002.

20-DEC-2001; 2001WO-US050459.  
20-DEC-2000; 2000US-0256868P.  
30-MAR-2001; 2001US-0280186P.  
01-MAY-2001; 2001US-0287735P.  
05-JUN-2001; 2001US-0295848P.  
25-JUN-2001; 2001US-0300465P.

(BRIM ) BRISTOL-MYERS SQUIBB CO.  
Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,  
Siemers N, Bol D, Schieven G, Finger J, Toddard CG, Bassolino D,  
Krystek S, Mcatee P, Suchard S, Banns D;  
WPI; 2002-599721/64.

Novel polynucleotides encoding human phosphatase polypeptides useful in the prevention or treatment of e.g. proliferative and cardiovascular disorders.

Disclosure; Fig 9; 80pp; English.

The invention relates to a novel isolated nucleic acid comprising a polynucleotide having a nucleotide sequence selected from 40 polynucleotides fully defined in the specification. The polynucleotide of the invention has antiproliferative, hepatotropic, nephrotropic, antiarthritic, antiproliferative, cardiac, and cytostatic activity. The polynucleotide may have a use in gene therapy. A polynucleotide or

polypeptide of the invention is useful for preventing, treating or ameliorating a medical condition, e.g. a proliferative disorder. They are also useful for treating e.g. liver disease, renal failure, immunological disorders including arthritis and psoriasis, cardiovascular disorders such as congenital heart defects and congestive heart failure, and cancer. A method of the invention is useful for diagnosing a pathological condition or susceptibility to a condition in a subject. The present sequence is used in the exemplification of the invention

Sequence 1705 AA;

Query Match 5.0%; Score 96.5; DB 5; Length 1705;  
Best Local Similarity 21.4%; Pred. No. 9.8;  
Matches 78; Conservative 40; Mismatches 134; Indels 113; Gaps 17;

2 RRGAG-----AARGRASCWALMLMLAVVPGW-----SRVSGIPSR----- 39  
32 RQGGGPPPLSVNVSRRKPT--SLPLSWAABDGGDYALCLRAMNLGFPBGQLOAHT 88  
40 -----HMPVPYKRPD-----FRPKDPYCOAKYFC-----PTGSP 71  
89 NSSFFHGLVPGSRKYQLVLR-----CWQNTVITLTARTAPVVRGLQHSIGSPA 143  
72 PV-----MEGDDIEVFRLQAP-----VMEFKYGDILGHLKTMHDAIGFR 111  
144 SLEASWSDASGDQSDYQLLHYHPSHTLACNVSVSPDTLSYNGDILPSQOYVLEVTWA 203  
112 STLTGKNYMEVLEQLNGCTEPHLPREMDAPFW-CNQACAFCEFGIDVHKENGTY 170  
204 GSHAKTSILQWTEPPYPHLLTRALGTSLOAFMNSSRGATWFHLITDLBGNILTKY 263  
171 QVATISGNMFMQAKKVKODNETGIY-----ETWVKASPEKGAETMFDSDYC 219  
264 VRGISTHFLRNS-----PGTPYQKICAAAGPHQWGNAT-----EWTYSPYS 310  
220 SKVLTETNKLAFFGAEFKNITNRYRIFLYSG---EPTLYGNE--TSVVG---PTGNKT 271  
311 DVLTLPLMELMWSWKGQARDGY--VTKLSGVVENTTTLGPECNAAVPPGLPPGHYT 368  
272 LGTAI 276  
369 LGLRV 373

RESULT 14  
ABB60074  
ID ABB60074 standard; protein; 2703 AA.

ABB60074;  
26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 7014.

Drosophila melanogaster polypeptide biology; cell signalling; insecticide; pharmaceutical.  
Drosophila melanogaster.  
WO200171042-A2.  
27-SEP-2001.  
23-MAR-2001; 2001WO-US009231.  
23-MAR-2000; 2000US-0191637P.  
11-JUL-2000; 2000US-00614150.  
(PEKE ) PE CORP NY.  
Venter JC, Adams M, Li PWD, Myers EW;  
WPI; 2001-656860/75.

N-PSDB; AB104177.  
New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.  
Disclosure; SEQ ID NO 7014; 21bp + Sequence Listing; English.  
The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB116173-AB116175). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences)  
Sequence 2703 AA;  
Query Match 5.0%; Score 95.5; DB 4; Length 2703;  
Best Local Similarity 20.1%; Pred. No. 24;  
Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;  
31 RVSGIPSRHWPVPRKPRPPYCOAKYFCPTGSPFPMWEGDDIVFRLOAFVW 90  
1986 RFNKIPQSEK-----KSKRPHPEIMWKKLEKT-----PYKGSNLAIPYQLEGINW 2032  
91 ERYGDIILGHLKIMDAIGFRSTLTGKNYTWMEYELFOLGNCCTPHEIPREMDAPFCNQG 15C  
2033 -LKSWMYTHNCIILADEWGLGKTIGLTFVHASYVEYGRG--PFLVIAPISTIPWQGE- 20E8  
151 AACFPEGIDVHWKENGTLVQVATISGKMNQAKWVQODNETGIYYETWNVKASPE--- 207  
2089 ---FEG-----WTDMNVVYVHGSVT-----SKQMIQDYE--YYXTESGKVLKEPIK 2130  
208 -KGAETWDSYDCSKFVLRTEFN-----KLAEFGEFKNIE----- 241  
2131 ENVLITTEMTVTDYMDKAFNMRLCVIDEAHLKRNKCKLLE-GLRQJINIEHRYLSGT 21E9  
242 ---TNYTRIF--LYSGEPTYLGNETSVEFGTGNKTLGLAIRFYYPFPHL----- 287  
2190 PLQNNISLFLSLNLFEPSSQSEFMSSEFSGSLRTEEVNKLQALLKPMRLRLKDVE 2249  
288 ----PTKFFLLSLQIFDAVIVHKOFYLFYNFETVFLPMKSPFIK--ITYEIP 335  
2250 KSLAPKEETIEV---ELTNIOKKYV-----RGLLEQNSFLKKGTTTSANIP 2293  
SQUIT 15  
3052308  
ABU52308 standard; protein; 429 AA.  
ABU52308;  
07-MAY-2003 (first entry)  
Helicobacter pylori selected interacting domain (SID) protein #1651.  
Protein-protein interaction; ulcer; selected interacting domain; SID.  
Helicobacter pylori.  
WO200266501-A2.  
29-AUG-2002.  
28-DEC-2001; 2001WO-EP015428.  
02-JAN-2001; 2001US-0259302P.  
(HYBR-) HYBRIGENICS.

PA (INSP ) INST PASTEUR.  
XX  
PI Legrain P, Rain J, Collard F, De Reuse H, Labigne A;  
XX WPI; 2002-674910/72.  
XX  
DR N-PSDB; ABX70383.  
XX  
PT New complexes of protein-protein interactions in Helicobacter pylori,  
PT useful for identifying modulating compounds for treating or preventing  
PT ulcers in mammals.  
PS  
XX  
XX Claim 6; Page 618; 642pp; English.  
XX  
CC The invention describes a complex of protein-protein interactions in  
CC Helicobacter pylori selected from 421 complexes given in the  
CC specification. The complex of protein-protein interactions are useful for  
CC screening for agents which modulate the interaction of proteins.  
CC Modulating compounds which binds to a targeted bacterial protein may be  
CC used for treating or preventing ulcers in a human or animal. This is the  
CC amino acid sequence of a selected interacting domain (SID), identified  
CC via protein-protein interactions. Note: Where the patent number printed  
CC at the top of the pages in the specification has obscured areas of  
CC protein sequence, the indexer has replaced the residue with an X to  
CC represent an illegible residue  
XX  
SQ Sequence 429 AA;

Query Match 4.9%; Score 94; DB 5; Length 429;  
Best Local Similarity 23.6%; Pred. No. 2.3;  
Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;  
QY 110 FRSTLTGKNYTWMEYELFOLGNCCTPHEIPREMDAPFCN-----QGAACFPEGIDVHWK 164  
DB 125 FTQYVGGKNSALV-----NATTP-----MANGSIPKSNSTVREGGIEGVWNG 167  
QY 165 ENGLVQVAT-----ISGNNFNQAKWVQODNETGIYYETWNVKASPEKGAETWDSYDC 219  
DB 168 KTGYITGFTFADRVYITGNMW-----TG-----NGAQTGGA--T 200  
QY 220 SKFVLRTFNKLAEGAFKNIET--NYTRIFYSGEPTYLGNETSVEFGTGNKTLGLAI 276  
DB 201 LNFVGATEINIA--GATFKMLKTSQNSYTFMALGD-----SSGSAKINVSQ 246  
QY 277 KREF-----YPFKPHPTKEFLLSLQIFDAVIVHKOFYLF-----YNFEYV-FLPMK 323  
DB 247 SDFQDMGGGVDYFGN-----GVFDSVFNKAYYFGQTEKSYNFKNTNFFLAGN 295  
QY 324 FPF-I-KITYERIPUPIRNKTLVG 345  
DB 296 FKFGKTTIEKSVLSDASYTFDG 318

Search completed: June 8, 2004, 14:03:55  
Job time : 94.7861 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 14:00:33 ; Search time 26.9111 Seconds

(without alignments)  
663.762 Million cell updates/sec

title: US-10-010-050a-2  
affect score: 1927  
sequence: 1 MRRGAGARGASWCWALAL.....IKITYEIRPLIRNKTLISGL 346

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

set-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	91	4.7	4563	4 US-09-108-006C-1	Sequence 1, Appli
2	88.5	4.6	1462	3 US-07-792-600-31	Sequence 31, Appl
3	88.5	4.6	1462	3 US-09-157-021-31	Sequence 31, Appl
4	88.5	4.6	1462	3 US-09-156-842-31	Sequence 31, Appl
5	88.5	4.6	1462	4 US-09-591-514-31	Sequence 31, Appl
6	88	4.6	608	4 US-09-489-039A-13503	Sequence 13503, A
7	87.5	4.5	178	4 US-09-134-000C-3624	Sequence 3624, Ap
8	87.5	4.5	876	4 US-09-254-352B-19	Sequence 19, Appl
9	83	4.3	555	4 US-09-134-078-24	Sequence 24, Appl
10	83	4.3	1056	4 US-09-079-030-217	Sequence 217, Appl
11	83	4.3	1310	3 US-08-989-299-10	Sequence 10, Appl
12	83	4.3	1310	4 US-09-407-422B-27	Sequence 10, Appl
13	83	4.3	4536	4 US-09-180-422B-27	Sequence 27, Appl
14	83	4.3	4536	4 US-09-079-030-1	Sequence 1, Appli
15	82	4.3	322	4 US-09-252-991A-26565	Sequence 26565, A
16	81	4.2	741	4 US-09-489-039A-7854	Sequence 7854, Ap
17	81	4.2	771	4 US-09-462-284-2	Sequence 2, Appli
18	80.5	4.2	718	4 US-09-328-352-4640	Sequence 4640, Ap
19	80	4.2	451	4 US-09-489-039A-11849	Sequence 11849, A
20	80	4.2	607	4 US-09-537-682-1	Sequence 1, Appli
21	80	4.2	652	4 US-09-489-039A-12461	Sequence 12461, A
22	79.5	4.1	197	4 US-09-489-039A-8317	Sequence 8317, Ap
23	79	4.1	844	4 US-09-564-805-227	Sequence 227, Appl
24	78.5	4.1	634	4 US-09-849-334-4	Sequence 4, Appli
25	78.5	4.1	634	4 US-10-274-878-4	Sequence 4, Appli
26	78.5	4.1	669	4 US-09-134-000C-6185	Sequence 6185, Ap
27	77.5	4.0	626	2 US-08-956-242-2	Sequence 2, Appli

28	77.5	4.0	626	3 US-09-351-215-2	Sequence 2, Appli
29	77	4.0	198	4 US-09-252-991A-19295	Sequence 19295, A
30	77	4.0	238	4 US-08-634-475-6	Sequence 6, Appli
31	77	4.0	238	4 US-09-709-791-6	Sequence 6, Appli
32	77	4.0	405	4 US-09-134-001C-5259	Sequence 5259, Ap
33	77	4.0	501	2 US-08-781-802-4	Sequence 4, Appli
34	77	4.0	501	2 US-08-781-802-6	Sequence 4, Appli
35	77	4.0	501	3 US-08-694-078-4	Sequence 4, Appli
36	77	4.0	501	3 US-08-694-078-6	Sequence 4, Appli
37	77	4.0	501	3 US-09-058-260-4	Sequence 4, Appli
38	77	4.0	501	3 US-09-058-260-6	Sequence 4, Appli
39	77	4.0	501	3 US-09-058-260-14	Sequence 14, Appli
40	77	4.0	501	3 US-09-058-260-20	Sequence 20, Appli
41	77	4.0	501	3 US-09-058-260-22	Sequence 22, Appli
42	77	4.0	501	3 US-09-058-260-24	Sequence 24, Appli
43	77	4.0	501	3 US-09-058-260-32	Sequence 32, Appli
44	77	4.0	615	4 US-09-107-532A-6507	Sequence 6507, Ap
45	77	4.0	619	4 US-09-107-532A-4897	Sequence 4897, Ap

## ALIGNMENTS

RESULT 1  
US-09-108-006C-1  
; Sequence 1, Application US/09108006C

; Patent No. 6524613  
GENERAL INFORMATION:

APPLICANT: Steer, Clifford J.

Kren, Betsy T.

Bandyopadhyay, Paramita

Roy-Chowdhury, Jayanta

TITLE OF INVENTION: Hepatocellular Chimera

NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:

ADDRESS: Kimeragen, Inc.

STREET: 300 Pheasant Run

CITY: Newtown

STATE: PA

COUNTRY: USA

ZIP: 18940

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09108, 006C

FILING DATE: 30-Jun-1992

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/054,288

FILING DATE: 30-Apr-1997

APPLICATION NUMBER: 60/054,837

FILING DATE: 05-Aug-1997

APPLICATION NUMBER: 60/064,996

FILING DATE: 10-Nov-1997

APPLICATION NUMBER: 60/074,497

FILING DATE: 12-Feb-1998

APPLICATION NUMBER: PCT US 98/08834

FILING DATE: 30-Apr-1998

ATTORNEY/AGENT INFORMATION:

NAME: Friebe, Thomas

REGISTRATION NUMBER: 29258

REFERENCE/DOCKET NUMBER: 7991-015-999

TELEPHONE: 215-504-4545

TELEFAX: 215-504-4444

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4563 amino acids

TYPE: amino acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 IS-09-108-006C-1

Query Match 4.7%; Score 91; DB 4; Length 4563;

Best Local Similarity 20.4%; Pred. No. 7.7;  
 Matches 58; Conservative 38; Mismatches 122; Indels 66; Gaps 11;

```

Y 74 MEQDDIEVERLQAPWEPKYDILG--HLKTMHDAIGRSTLTGKNYMEVELFQUG 130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4013 MDEDDPFS-----KMFYFSPQSPDKKLTIFKTELRYRESDEDTQYKVNNEEASAG 4065
Y 131 NCFPHLRPEMDAPFMCNOGACFEBSIDVHKENS--TLVQVATISGMNFQMAKVMQ 189
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4066 LLT--SLKDVV-----KATGVLYDVNKKYHWHBHGTLREVSCKRRNLQCHAEVYQ 4117
Y 190 DNETGIYETWNVKASPEKAEFTWFSYDCKEVLRTFNKLAEFGEFKNIETN--YTRIF 248
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4118 ----GAIRELDDIDERQKASGTTGYQ-----EMKQAKNLVQELL 4156
Y 249 LYSGEPTLYGNETSVPGPTGNKTYGLAIK-----RYYFPKPHLPYKEFL 293
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4157 TOEGQASFOGKQNVFDGLVAVTQDEFHMKVYKHLIDSLIDFNPRFQFPKPGIYREEL 4216
Y 294 LSLIQIEDAVIYHKOY-----LFNFEYWFLLMKKEPKIK 328
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 4217 CTMF--IREVGTLISQVYSKYHNGSEILFSYFQDLVITLPELAK 4259

```

# RESULT 2

IS-07-792-600-31  
 Sequence 31, Application US/07792600  
 Patent No. 6008045

## GENERAL INFORMATION:

APPLICANT: COPELAND, WILLIAM C.  
 APPLICANT: WANG, TERESA S.-F.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
 TITLE OF INVENTION: TEMPLATE-DEPENDENT ENZYMAITC SYNTHESIS OF NUCLEIC ACID  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Peter G. Carroll  
 STREET: 220 Montgomery Street, Suite 710  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/792,600  
 FILING DATE: 19911115  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CARROLL, PETER G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: STDU-00097  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1462 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 IS-07-792-600-31

Query Match 4.6%; Score 88.5; DB 3; Length 1462;

Best Local Similarity 20.3%; Pred. No. 2.5;  
 Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

```

QY 46 KRFDFPKDPYQQAQKTYFCPTG-----SPIPWEGD---DDIEV---FRL 85
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 183 KRSISASNPSPVHTATVPSGKIASPVSRKEPPLTPVLKRAEPAGDVQVESTEEBQ 242
QY 86 QAPWFEKKDGLGHKIMHDAIGFRSTLTGKNYMEVELFQUGNCTFPHLRPEMDAPF 145
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 243 ESGAMEFEDGDF-----DERNEVEV-----DLEP--MAAKA 272
QY 146 WCNQGAACFEBSIDVHKEN--GLTVQVATISGMNFQMAK--VKQDNETGIYETWN 201
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 273 WDKES-----EPAEVYKQADSGKTV-----SYLGSFLPDVSCMDIDQEDGSFVSQEVQ 323
QY 202 VKASP---EKGAET-----WDSYD---CSFVYRTFNKLAEPGAE-----FKNI 240
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 324 VDSHLLPLVKGADBEQVFFHYWLDAVEDQYNOGVVFLGKWAIESAEHTVSCCVWKN 383
QY 241 ETNYTRIFLYSGEPTLYGNETSVPGPTGNKT--LGLAIKRFYFPKPHLPYKEFLSLIQI 299
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 384 ERT-----LY-----FLPRMKIDLTNGKETGPISMKVYEEFDEKIAIKYIMK---- 429
QY 300 FPAVYHKOYLFYNEFYWFLPMKPPKIITYE---ELPPLRNKTLG 344
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 430 FSKSPYEKN---YAFELPDVPEKSEYLEVKYSAEMPQLPOLDKGETFS 474

```

# RESULT 3

US-09-157-021-31  
 Sequence 31, Application US/09157021A  
 Patent No. 6100023

## GENERAL INFORMATION:

APPLICANT: Copeland, William C.  
 APPLICANT: Wang, Teresa S. F.  
 TITLE OF INVENTION: Drug Design Assay  
 FILE REFERENCE: STDU-03484  
 CURRENT APPLICATION NUMBER: US/09/157,021A  
 EARLIER FILING DATE: 1998-09-18  
 EARLIER APPLICATION NUMBER: 07/792,600  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: Patent Ver. 2.0  
 SEQ ID NO 31  
 LENGTH: 1462  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-157-021-31

Query Match 4.6%; Score 88.5; DB 3; Length 1462;

Best Local Similarity 20.3%; Pred. No. 2.5;  
 Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

```

QY 46 KRFDFPKDPYQQAQKTYFCPTG-----SPIPWEGD---DDIEV---FRL 85
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 183 KRSISASNPSPVHTATVPSGKIASPVSRKEPPLTPVLKRAEPAGDVQVESTEEBQ 242
QY 86 QAPWFEKKDGLGHKIMHDAIGFRSTLTGKNYMEVELFQUGNCTFPHLRPEMDAPF 145
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 243 ESGAMEFEDGDF-----DERNEVEV-----DLEP--MAAKA 272
QY 146 WCNQGAACFEBSIDVHKEN--GLTVQVATISGMNFQMAK--VKQDNETGIYETWN 201
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 273 WDKES-----EPAEVYKQADSGKTV-----SYLGSFLPDVSCMDIDQEDGSFVSQEVQ 323
QY 202 VKASP---EKGAET-----WDSYD---CSFVYRTFNKLAEPGAE-----FKNI 240
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 324 VDSHLLPLVKGADBEQVFFHYWLDAVEDQYNOGVVFLGKWAIESAEHTVSCCVWKN 383
QY 241 ETNYTRIFLYSGEPTLYGNETSVPGPTGNKT--LGLAIKRFYFPKPHLPYKEFLSLIQI 299
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D 384 ERT-----LY-----FLPRMKIDLTNGKETGPISMKVYEEFDEKIAIKYIMK---- 429

```

[illegible][illegible]

Y 231 -----AEF-----GAEFKNIETNYTRIFLYSGEP-TYLGNETSVFG---PTGNKTLGLAIK 277  
b 448 DSHDTARFKSLDKVDVARLPLAVVWLFPMGVGPCIYYGGEVGVGDNNDPCKRP----- 501  
Y 278 RFFYPFKPHLPTEKFELISLIQ 298  
b 502 ---FPWDPALQDTQ-LLAAYQ 518

RESULT 7  
S-09-134-000C-3624  
Sequence 3624, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3624  
LENGTH: 178  
TYPE: PRT  
ORGANISM: Enterococcus faecalis

S-09-134-000C-3624

Query Match 4.5%; Score 87.5; DB 4; Length 178;  
Best Local Similarity 21.6%; Pred. No. 0.12; Matches 41; Conservative 24; Mismatches 60; Indels 65; Gaps 3;

Y 117 KNYTMEWELFQJGNCPTPLRPMDAPPMGCGAACFEGCIDDVHKENGTIVQVATIS 175  
b 13 KNYTRLGRITDLSKGIISQLERDLSP-----SWETFF-----SILEVLGYT 55  
Y 177 GNNENGM-----AKWYQDNETGYVE-----TNVVKASPEKGATWFDSDCSFYVR 225  
b 56 PROFFSQQPLEQKLYNEBSSTLYDEHGVELKMLIPASNEKEMP-----VII 105  
Y 226 TENKLAEPGAEPKNIETNYTRIFLY-----SGEPTYLGNETSVFGPTGNKTLGLAI 275  
b 106 TPDK-----NGEYKTEPSPSETPFIYIDSIRLTGEAAYIAKK-----GQT----- 143  
Y 277 KREYYPFKPH 286  
b 149 -WYQATEPH 157

RESULT 8  
S-09-254-352B-19  
Sequence 19, Application US/09254352B  
Patent No. 6365350  
GENERAL INFORMATION:  
APPLICANT: HAYASHIZAKI, Yoshinide  
TITLE OF INVENTION: METHOD OF DNA SEQUENCING  
FILE REFERENCE: 024705-080  
CURRENT APPLICATION NUMBER: US/09/254,352B  
PRIOR FILING DATE: 1999-11-10  
PRIOR APPLICATION NUMBER: PCT/JP98/03039  
PRIOR FILING DATE: 1998-07-06  
PRIOR APPLICATION NUMBER: JP 10-155847  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: JP 9-196478  
PRIOR FILING DATE: 1997-07-07  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 876  
TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T7 RNA polymerase  
US-09-254-352B-19

Query Match 4.5%; Score 87.5; DB 4; Length 876;  
Best Local Similarity 21.6%; Pred. No. 1.4;  
Matches 63; Conservative 28; Mismatches 107; Indels 93; Gaps 17;

Y 1 MRGAGARGASW---CWALLMLAVPG--WS-----RVSGLPSRR-----HMPVY 45  
b 267 ISKRAGALAGISPMHQPCCVPPVWETVGGYWSGVALVTHSKALRYADVHMPVY 326  
Y 46 KRDFRPRKDPYCOAK-----YTPCPGSPDPVMEGD-----DIEVFRQAQVY 89  
b 327 KAVNLAQNTPMKNNKVLAVNVEIYNMGCPYGD--VPALERELPRPPDIDTNEVARXA 385  
Y 90 WEKFGDLGHUKM-----HDAIGRSTLTGKNYMEW-----YELF 127  
b 386 WEKEAAAVRKDKARQSRRCCEFMVAQMFANHKAIME-----PYNMDWRGRVYAVS 438  
Y 128 QJGN-----CTPPIHP-EMDAPFMCN--OGACFEGCIDDVHKENGTIVQVATISGN 178  
b 439 MFGNDMTKSLITLAKKPTIGLDGFYWLKIHGANC--AGVDVPPPE-----RIKFIEN 491  
Y 179 MFWOMAK-----WYQDNETGIYETMNVKASPEKGATWFDSDCS 220  
b 492 EGNILASADPLNNTWMTQODSPFAFCFPAVAK-----HGNLSYNS 535

RESULT 9  
US-09-134-078-24  
Sequence 24, Application US/09134078  
Patent No. 6368844  
GENERAL INFORMATION:  
APPLICANT: Bylina, Edward J.  
TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/134,078  
FILING DATE: 13-AUG-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/949,026  
FILING DATE: 10-OCT-1997  
APPLICATION NUMBER: 60/056,916  
FILING DATE: 06-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Hailie, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/024002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 555 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal

S-09-134-078-24

Query Match 4.3%; Score 83; DB 4; Length 555;

Best Local Similarity 22.1%; Pred. No. 2,3; Mismatches 50; Conservative 32; Indels 68; Gaps 12;

```
Y 64 FCPGSPIMWEGDDIEVRLQAPWPEFYKGLHLKIMDAIGFRSLTGKNTMWM 123
b 149 FVPL-EPVLVLEDPN-----TPLLREKVALVC---MNNARPKHTPTG---WCSM 193
Y 124 YEAF--QLGNCFTPHLRPMMDAFWCONGACFFEGIDVHWK-----164
b 194 YHYFLDLTWEETLXNKJAKNAFFEFVFO-----IDDAVEKDIGDMLVTRGDFPSVE 243
Y 165 -----ENGLVQVATI-----SGMFMQMAKM-VKQDNNGCIYVETNVAASPEKGA 211
c 246 MAKVIAENGEFPGIWTAPPSVSSTSDVFNHDPDVVKENGEPKMAFRNNKK-----297
f 212 TWFDSDYDCK-----FVLRFTFNKLAFFGAFKNIETNYRIPLYSQ 252
c 298 --IYALDSLKDEVLNMLFDLFSRLRMGYRYFKID-----FLFAG 335
```

RESULT 10

3-09-079-030-217

Sequence 217, Application US/09079030

Patent No. 6635623

GENERAL INFORMATION:

APPLICANT: Guevera, Jr., Juan G.

APPLICANT: Hoogeveen, Ron C.

APPLICANT: Moore, Paul J.

TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY

TITLE OF INVENTION: VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS

NUMBER OF SEQUENCES: 229

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White &amp; Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,030

FILING DATE: Concurrently Herewith

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: McMillian, Nabeela R.

REGISTRATION NUMBER: P-43,363

REFERENCE/DOCKET NUMBER: ARAG:003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 217:

SEQUENCE CHARACTERISTICS:

LENGTH: 1056 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

3-09-079-030-217

Query Match 4.3%; Score 83; DB 4; Length 1056;

Best Local Similarity 19.7%; Pred. No. 6,3;

Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

```
74 MEGDDIVFRLQAPWPEFYKGLHLKIMDAIGFRSLTGKNTMWMYELFQLG 130
506 WDEDDFS-----KMFYSPSSPPKLTITFTLVRBSDETOIKVWEESAAG 556
```

QY 131 NCTPHLRPEMDAPFWCONGACFFEGIDVHWKENG-TLVQVATISGMFMQMAKWKQ 189

Db 559 LIT--SLKRNVP-----KATGVLIDYVKKYHMTGLTLREVSGLRRLNLAEMWVQ 610

QY 190 DNETGIYETNVAASPEKGAETWSDYDCKFVLRTFNKLAFFGAFKNIETN-YTRIF 248

Db 611 ---GAIQIDIDIVRFGKASGTGTQ-----EMDXQNLVQELL 649

QY 249 LYSGEPTIGNFTSVFGPTGNKTLGLAK-----RIFYPKPLPYKEFL 293

Db 650 TOEGQASEFGLKDNVFDGLVRYQKFMKVKHLIDSLIDFLNFPFQFGKPGIYREEL 709

QY 294 LSLIDPAVIVHKOY-----LFNFEYVFLPMKEPPIK 328

Db 710 CTMF-IREYGVLSQVYSKVNHSSEILFSYFODLVITLPELRK 752

RESULT 11

US-08-989-299-10

Sequence 10, Application US/08989299

Patent No. 6194556

GENERAL INFORMATION:

APPLICANT: Acton, Susan L.

APPLICANT: Robinson, Keith E.

TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG

TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG &amp; ELIOT LLP

STREET: One Post Office Square

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/989,299

FILING DATE: 11-DEC-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Arnold B., Beth

REGISTRATION NUMBER: 35,430

REFERENCE/DOCKET NUMBER: MIA-025.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-832-1000

TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1310 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-989-299-10

Query Match 4.3%; Score 83; DB 3; Length 1310;

Best Local Similarity 19.7%; Pred. No. 8,8;

Matches 77; Conservative 61; Mismatches 145; Indels 108; Gaps 23;

```
QY 15 CWAL-----ALIMIAVPGMSRVSGIPSRHWPVYKRFDFRPKDPYCOA 60
Db 170 CWSLDPLDNLNIIASSRSYAML-LFAMEGMHNAVGIPLK---PL-YOEFYALSN-BAYROD 223
QY 61 KYTCTPGS-----PIPWEGDDIE-VFRLQAPWPEFYKGLHL-KIMDAIGFR- 111
Db 224 --GSDIDAYKRSWYDSPTFE--EDLERIYHOLEPL---YLNLAAYRRVLAHRRYGRY 275
QY 112 -----STLTGKNTMWMYELFQLGNCFTPHLRPEMDAPFWCONGACFFEGIDVHW 163
```

b 276 INRGPIPAHLGNMMAQSWESIYDM-VVFPD-KPULDV----- 313  
Y 164 KENGLIVQATTISGNWENQAKWQKQDNERTGIYETW--NVKASPEKG-----AETWPD 215  
b 314 --TSTWQKGMNATHMRVVAEEFFTSIGLLPMPPEFAESMLEKPEDGRVAVCHASAM-D 370  
Y 216 SYDQSKFVLTFFNKLAFFGAEPKNIETNYRIFL-VSGEPYVLTGNETSVEPGTKNTLGL 274  
b 371 FVNRKDFRIKQCTQVMTDQSTVHHEMGHVQYLLQYKQDQVSLRANPGFHEAIGVIAL 430  
Y 275 AIKRFYFFPHL-----PTKEFLSLLOI-----FDVAVIHKQFYL----- 312  
b 431 SVSTPAHLHRTIGLDHVNTDESINYLKMALEKIAFLPFGLVQWQWGVSGRTPSS 490  
Y 313 -YNEFWFLPMKFPFKITYEELPLPRNKT 342  
b 491 RYNEFWMLRTKYQGI-----CPVVRNET 515

## ESTUT 12

S-09-407-427-10

Sequence 10, Application US/09407427

Patent No. 6610497

## GENERAL INFORMATION:

APPLICANT: Acton, Susan J.

APPLICANT: Robison, Keith E.

TITLE OF INVENTION: ANGIOGENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC

FILE REFERENCE: MNT-132CP2

CURRENT APPLICATION NUMBER: US/09/407,427

CURRENT FILING DATE: 1999-09-29

PRIOR APPLICATION NUMBER: 09/163,648

PRIOR FILING DATE: 1998-09-30

PRIOR APPLICATION NUMBER: 08/989,299

PRIOR FILING DATE: 1997-12-11

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 10

LENGTH: 1310

TYPE: PR1

ORGANISM: Oryctolagus cuniculus

S-09-407-427-10

Query Match 4.3%; Score 83; DB 4; Length 1310;

Best Local Similarity 19.7%; Pred. No. 8.8; Matches 77; Conservative 61; Mismatches 145; Indels 108; Gaps 23;

Y 15 CVAL-----ALLMLAVVPGMSVSGISRRHVPVYKRFPRPRPYCOA 60  
b 170 CWSIDPDIANNILASSRSYAML-LFAWEGMNAVGIPLK--PL-YOEFALSN-EAYROD 223  
Y 61 KYTFCPTGS-----PIPVNEGDDIE-VRLOAPVWEFKYGDLLGH-LKIMMDALGFR- 111  
b 224 --GSDTGAARWSWYDSPTFE--EDLERIYHQLEPL--YINLHVVRVLIHRRYGDRT 275  
Y 112 -----STLTGKNYTWMEYELFOLGNCCTPHLRPEMDAPFWQNOGAACFFEGIDVHW 163  
b 276 INRGPIPAHLGNMMAQSWESIYDM-VVFPD-KPULDV----- 313  
Y 164 KENGLIVQATTISGNWENQAKWQKQDNERTGIYETW--NVKASPEKG-----AETWPD 215  
b 314 --TSTWQKGMNATHMRVVAEEFFTSIGLLPMPPEFAESMLEKPEDGRVAVCHASAM-D 370  
Y 216 SYDQSKFVLTFFNKLAFFGAEPKNIETNYRIFL-VSGEPYVLTGNETSVEPGTKNTLGL 274  
b 371 FVNRKDFRIKQCTQVMTDQSTVHHEMGHVQYLLQYKQDQVSLRANPGFHEAIGVIAL 430  
Y 275 AIKRFYFFPHL-----PTKEFLSLLOI-----FDVAVIHKQFYL----- 312  
b 431 SVSTPAHLHRTIGLDHVNTDESINYLKMALEKIAFLPFGLVQWQWGVSGRTPSS 490  
Y 313 -YNEFWFLPMKFPFKITYEELPLPRNKT 342

Db 491 RYNEFWMLRTKYQGI-----CPVVRNET 515

## RESULT 13

US-09-180-422B-27

Sequence 27, Application US/09180422B

Patent No. 6444644

## GENERAL INFORMATION:

APPLICANT: BRUCKDORFER, KARL R

ETTELBAIE, CAMILIE

TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED

FROM APOLIPOPROTEIN B-100

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON &amp; VANDERHEYE, P.C.

STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/180,422B

FILING DATE: 07-Dec-1998

CLASSIFICATION: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36663

REFERENCE/DOCKET NUMBER: 117-268

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164000

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 4536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 27:

US-09-180-422B-27

Query Match 4.3%; Score 83; DB 4; Length 4536;

Best Local Similarity 19.7%; Pred. No. 63; Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

QY 74 MEGDDDIEVFRLOAPVWEFKYGDLLG--HLKIMMDALGFRSTLTGKNYTWMEYELFOLG 130  
Db 3986 KQEDDDFS-----KKNFYSPQSSPDKLTITFTLELVRESDEDTQIKVNEEBAAG 4038  
QY 131 NCTFPHLRPEMDAPFWQNOGAACFFEGIDVHWKENG-TLVQVATTISGNWENQAKWQK 189  
Db 4039 LFL--SLKQNVF-----KATGVLYDVWKYVMEHTGLTLREVSRLRNLTQNNMAEYVQ 4090  
QY 190 DNETGIYETWAVKASPEKGACTWPDSDYCSKFLVLTENKLAFFGAEPKNIETN-YTRIF 248  
Db 4091 ---GAIRQIDIDIVRFQKAAAGTTGTQY-----EMWDRKQNLVQEL 4129  
QY 249 LYSGEPTVLTGNETSVEPGTKNTLGLAIF-----REYVFPKPLPTKEFL 293  
Db 4130 TQEGQASFGGLKQNDVFDGLVRVTOQKHKVYKHLISLDLFLNFPKQFQKRGITREEL 4189  
QY 294 LSLQIFDAVIVHKQFY-----LFYNEFWFLPMKFPFKIT 328  
Db 4190 CTFM-FIREVGTVLSQYVSKVHNGSEILFSYFQDIVITLPEFLRK 4232

## RESULT 14

US-09-079-030-1





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 14:04:04 ; Search time 71.1222 Seconds

(without alignments)  
1368.675 Million cell updates/sec

File: US-10-010-050a-2

Effect score: 1927  
Sequence: 1 MRGAGARGRASWCMALAL.....IKTYEELPLPRNKTLSGL 346

oring table: BLOSUM62

arched: Gapop 10.0 , Gapext 0.5

total number of hits satisfying chosen parameters: 1155919

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppa/PCF\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppa/CTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppa/US10A\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

sult No.	Score	Query Match	Length	ID	Description
1	1927	100.0	346	US-09-122-383-2	Sequence 2, Appl1
2	1927	100.0	346	US-10-010-050a-2	Sequence 2, Appl1
3	1916	99.4	346	US-10-653-595-126	Sequence 126, Appl
4	1916	99.4	347	US-09-397-945-126	Sequence 126, Appl
5	886	46.0	163	US-09-864-761-39378	Sequence 39378, A
6	472	24.5	89	US-10-264-049-4034	Sequence 4034, Ap
7	97.5	5.1	328	US-10-369-493-12714	Sequence 12714, A
8	96	5.0	290	US-10-369-493-3172	Sequence 3172, Ap
9	95.5	5.0	836	US-10-408-765a-58	Sequence 58, Appl
10	94	4.9	1238	US-10-282-122a-58749	Sequence 58749, A
11	93	4.8	478	US-10-282-122a-55615	Sequence 55615, A
12	91.5	4.7	347	US-10-424-599-262855	Sequence 262855, A
13	91.5	4.7	350	US-10-425-114-57996	Sequence 57996, A
14	91	4.7	742	US-09-374-046a-160	Sequence 160, App
15	91	4.7	742	US-10-616-263-160	Sequence 160, App

16	91	4.7	742	US-10-408-765a-2332	Sequence 2332, Ap
17	91	4.7	4563	US-09-870-759-128	Sequence 128, App
18	91	4.7	4563	US-09-751-708a-128	Sequence 128, App
19	89	4.6	495	US-10-424-599-252265	Sequence 252265, A
20	89	4.6	567	US-10-133-797-4	Sequence 4, Appl1
21	88	4.6	419	US-10-028-245-2	Sequence 2, Appl1
22	88	4.6	605	US-10-282-122a-60098	Sequence 60098, A
23	88	4.6	3194	US-10-282-122a-59046	Sequence 59046, A
24	87	4.5	1160	US-10-369-493-12860	Sequence 12860, A
25	86	4.5	411	US-10-282-122a-60577	Sequence 60577, A
26	85.5	4.4	483	US-10-425-114-43218	Sequence 43218, A
27	85.5	4.4	483	US-10-425-114-43218	Sequence 43218, A
28	84.5	4.4	224	US-10-461-990-35	Sequence 35, Appl
29	84.5	4.4	513	US-10-425-114-46548	Sequence 46548, A
30	84.5	4.4	513	US-10-425-114-66071	Sequence 66071, A
31	84	4.4	343	US-10-425-114-45828	Sequence 45828, A
32	84	4.4	356	US-10-424-599-225305	Sequence 225305, A
33	84	4.4	719	US-10-425-114-56258	Sequence 56258, A
34	83.5	4.3	373	US-10-012-819-198	Sequence 198, App
35	83.5	4.3	434	US-10-084-749-2336	Sequence 2336, App
36	83.5	4.3	984	US-10-369-493-8357	Sequence 8357, Ap
37	83	4.3	552	US-09-909-464a-2	Sequence 2, Appl1
38	83	4.3	555	US-10-121-032-24	Sequence 24, Appl
39	83	4.3	555	US-10-093-037-24	Sequence 24, Appl
40	83	4.3	1086	US-10-369-493-5963	Sequence 5963, Ap
41	83	4.3	4563	US-09-802-640-32	Sequence 32, Appl
42	83	4.3	4563	US-10-403-902a-32	Sequence 32, Appl
43	82.5	4.3	381	US-10-032-585-7666	Sequence 7666, Ap
44	82.5	4.3	591	US-10-081-872-156	Sequence 156, App
45	82.5	4.3	591	US-10-385-305-156	Sequence 156, App

## ALIGNMENTS

RESULT 1  
US-09-122-383-2  
Sequence 2, Application US/09122383a  
Patent No. US20020042093A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
TITLE OF INVENTION: CHROMOSOME 13  
FILE REFERENCE: 97-38  
CURRENT APPLICATION NUMBER: US/09/122,383A  
CURRENT FILING DATE: 1998-07-24  
EARLIER APPLICATION NUMBER: 60/053,613  
EARLIER FILING DATE: 1997-07-24  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-122-383-2

Query Match	100.0%	Score 1927	DB 9	Length 346
Best Local Similarity	100.0%	Pred. No. 6e-181		
Matches 346	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MRGAGARGRASWCMALALIMLAVVGVGSRVSGIPSRHMPVYKRFDFPKDPYCOA	60	
DB	1	MRGAGARGRASWCMALALIMLAVVGVGSRVSGIPSRHMPVYKRFDFPKDPYCOA	60	
QY	61	KTYFCPTGSPIPVWEGDDIEVEFLQAPVWEFKYGDILGHLKIMHDAIGFSTLTGKNYT	120	
DB	61	KTYFCPTGSPIPVWEGDDIEVEFLQAPVWEFKYGDILGHLKIMHDAIGFSTLTGKNYT	120	
QY	121	MEWEFLFOLNCCTFPHLRPEMDAPFVNCQACFFEGIDVHWKENTGVQVATISGMF	180	
DB	121	MEWEFLFOLNCCTFPHLRPEMDAPFVNCQACFFEGIDVHWKENTGVQVATISGMF	180	

```

Y 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
Y 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
Y 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
Y 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
Y 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|
Y 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|

```

# RESULT 2

Sequence 2, Application US/10010050A  
Publication No. US20020173624A1

## GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Gilbertson, Debra G.

TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

TITLE OF INVENTION: CHROMOSOME 13

FILE REFERENCE: 97-38C1

CURRENT APPLICATION NUMBER: US/10/010,050A

CURRENT FILING DATE: 2002-03-26

PRIOR APPLICATION NUMBER: US 09/122,383

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/053,613

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 2

LENGTH: 346

TYPE: PRT

ORGANISM: Homo sapien

S-10-010-050A-2

Query Match 100.0%; Score 1927; DB 13; Length 346;  
Best Local Similarity 100.0%; Pred. No. 66-181;

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y 1 MRGAGAAAGRASWCALALMLAVVPGMSRVSGIPSRHWPVYKRFPRPDPYCOA 60
|
|
|
Y 1 MRGAGAAAGRASWCALALMLAVVPGMSRVSGIPSRHWPVYKRFPRPDPYCOA 60
|
|
|
Y 61 KYTFCPTGSPIPVMEGDDIEVFRLQAPVWEFRKYGDLGLKIMHDAIGFRSTLTGKNT 120
|
|
|
Y 61 KYTFCPTGSPIPVMEGDDIEVFRLQAPVWEFRKYGDLGLKIMHDAIGFRSTLTGKNT 120
|
|
|
Y 121 MEWEELFQIGNCTFPHLRPEMDAPFWCNGAACFFEGIDVHMKENGTLVQVATISGNMF 180
|
|
|
Y 121 MEWEELFQIGNCTFPHLRPEMDAPFWCNGAACFFEGIDVHMKENGTLVQVATISGNMF 180
|
|
|
Y 121 MEWEELFQIGNCTFPHLRPEMDAPFWCNGAACFFEGIDVHMKENGTLVQVATISGNMF 180
|
|
|
Y 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
Y 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
Y 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
Y 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
Y 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|
Y 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|

```

# RESULT 3

S-10-653-595-126

Sequence 126, Application US/10653595

Publication No. US20040048304A1

## GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

```

; FILE REFERENCE: P2027P1C1
; CURRENT APPLICATION NUMBER: US/10/653,595
; CURRENT FILING DATE: 2003-09-03
; PRIOR APPLICATION NUMBER: US 09/397945
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (242)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (246)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-653-595-126

```

Query Match 99.4%; Score 1916; DB 12; Length 346;  
Best Local Similarity 99.4%; Pred. No. 7,36-180;

Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MRGAGAAAGRASWCALALMLAVVPGMSRVSGIPSRHWPVYKRFPRPDPYCOA 60
|
|
|
DB 1 MRGAGAAAGRASWCALALMLAVVPGMSRVSGIPSRHWPVYKRFPRPDPYCOA 60
|
|
|
QY 61 KYTFCPTGSPIPVMEGDDIEVFRLQAPVWEFRKYGDLGLKIMHDAIGFRSTLTGKNT 120
|
|
|
DB 61 KYTFCPTGSPIPVMEGDDIEVFRLQAPVWEFRKYGDLGLKIMHDAIGFRSTLTGKNT 120
|
|
|
QY 121 MEWEELFQIGNCTFPHLRPEMDAPFWCNGAACFFEGIDVHMKENGTLVQVATISGNMF 180
|
|
|
DB 121 MEWEELFQIGNCTFPHLRPEMDAPFWCNGAACFFEGIDVHMKENGTLVQVATISGNMF 180
|
|
|
QY 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
DB 181 NQAKWVQDNETGTYETWNVKASPEKGAETWFDSDSCFVLRITNKLAEGFAEFKNI 240
|
|
|
QY 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
DB 241 ETNTRIFLYSGEPTLYGNETSVPFGTKNLGLAIKRFYFPKPHLPTEFELSLQIF 300
|
|
|
QY 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|
DB 301 DAVIVHKQFYLFYNFYFWFLPMKFPFIKITYEELPPIRNKTLISGL 346
|
|
|

```

# RESULT 4

US-09-397-945-126

Sequence 126, Application US/09397945

Publication No. US20030065139A1

## GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc. et al.  
TITLE OF INVENTION: 95 Human secreted proteins  
FILE REFERENCE: P2027PI  
CURRENT APPLICATION NUMBER: US/09/397,945  
CURRENT FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804  
PRIOR FILING DATE: 1999-03-18  
PRIOR APPLICATION NUMBER: 60/078,566  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,576  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,573  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,574  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,579  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,314  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080,312  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/078,578  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,581  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,577  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,563  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,313  
PRIOR FILING DATE: 1998-04-01  
NUMBER OF SEQ ID NOS: 470  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 126  
LENGTH: 347

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (242)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (246)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (347)

OTHER INFORMATION: Xaa equals stop translation

Query Match 99.4%; Score 1916; DB 12; Length 347;  
Best Local Similarity 99.4%; Pred. No. 7.3e-180;  
Matches 344; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MRGAGARAGRAACWALALMLAVVGVGSRVSGIPSRBRMPVYKRPDRPDDPCQA 60  
2 MRGAGARAGRAACWALALMLAVVGVGSRVSGIPSRBRMPVYKRPDRPDDPCQA 60  
61 KYTFCPTGSPFIPWEGDDIEVERLQAPVWEFKYGLGLKMLMDAIGRSTLTGKNT 120  
121 MEWYELFOAGNCFFPHLRPMMDAPFWCNOGAACFFEGIDVHMKENGTLVQVATISGNMF 180  
121 MEWYELFOAGNCFFPHLRPMMDAPFWCNOGAACFFEGIDVHMKENGTLVQVATISGNMF 180  
181 NQAKWVKQDNETGTYETWNVKASPEKGAETWEDSDSKFVLRTFNKLAERGAEPKXI 240  
181 NQAKWVKQDNETGTYETWNVKASPEKGAETWEDSDSKFVLRTFNKLAERGAEPKXI 240  
241 ETVYTRIFLYSGEFTYLGNETSVFGPTGNKTLGLAIKRFYEPKPLPTKEFLSLQIF 300  
241 ETVYTRIFLYSGEFTYLGNETSVFGPTGNKTLGLAIKRFYEPKPLPTKEFLSLQIF 300

QY 301 DAVYHKQFYLFYNEFWFLPMKPEFFIKITYEELPLPRNNTLSGL 346  
Db 301 DAVYHKQFYLFYNEFWFLPMKPEFFIKITYEELPLPRNNTLSGL 346

RESULT 5  
US-09-864-761-39378

Sequence 39378, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmca-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 39378

LENGTH: 163

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC001226.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

OTHER INFORMATION: EST\_HUMAN HTT: H01255.1, EVALUATE 5.00e-59

OTHER INFORMATION: SWISSPROT HIT: O75503, EVALUE 2.00e-96  
US-09-864-761-39378

Query Match 46.0%; Score 886; DB 9; Length 163;  
Best Local Similarity 100.0%; Pred. No. 6.5e-79;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 184 AKWVKQNEGTGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNKLAERGAEFKNIEYN 243  
1 AKWVKQNEGTGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNKLAERGAEFKNIEYN 60

Y 244 YTHIFLVSGBPVLGNSTVFGPNTKTLGLAIKRFYPPKPHLPTEKFLSLQIFDAV 303  
b 61 YTHIFLVSGBPVLGNSTVFGPNTKTLGLAIKRFYPPKPHLPTEKFLSLQIFDAV 120

Y 304 IVHKQFLFYFNFYFVLPKPFPIKITTYEIPLPINRKTLSGL 346  
b 121 IVHKQFLFYFNFYFVLPKPFPIKITTYEIPLPINRKTLSGL 163

RESULT 6  
S-10-264-049-4034

Sequence 4034, Application US/10264049  
Publication No. US20040005579A1  
GENERAL INFORMATION:

APPLICANT: Birse et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA133P1

CURRENT APPLICATION NUMBER: US/10/264,049

PRIOR FILING DATE: 2002-10-04

PRIOR APPLICATION NUMBER: PCT/US01/18569

PRIOR FILING DATE: 2001-06-07

PRIOR APPLICATION NUMBER: US 60/209,467

NUMBER OF SEQ ID NOS: 4360

SOFTWARE: PatentIn Ver. 3.1

SEQ ID NO 4034

LENGTH: 89

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (58)\_"

OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
S-10-264-049-4034

Query Match 24.5%; Score 472; DB 15; Length 89;  
Best Local Similarity 98.9%; Pred. No. 1.6e-38;  
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

258 GNETSVFSGPTGKTLGLAIKRFYPPKPHLPTEKFLSLQIFDAVIVHKQFLFYFNFY 317

1 GNETSVFSGPTGKTLGLAIKRFYPPKPHLPTEKFLSLQIFDAVIVHKQFLFYFNFY 60

318 WFLPKKFPPIKITTYEIPLPINRKTLSGL 346  
61 WFLPKKFPPIKITTYEIPLPINRKTLSGL 89

SUITE 7  
S-10-369-493-12714

Sequence 12714, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 12714

LENGTH: 328

TYPE: PRT

ORGANISM: Aspergillus nidulans

US-10-369-493-12714

Query Match 5.1%; Score 97.5; DB 15; Length 328;  
Best Local Similarity 26.1%; Pred. No. 0.69;  
Matches 41; Conservative 19; Mismatches 50; Indels 47; Gaps 7;

QY 168 TLVQVATLISGNMFNOMN-----KWKQDNETGYETWNNVAKASPEKGAET 212  
DB 57 TLVSGFGSNTHTQVAAVARKLGLVALVQEKWVMDKNGVYDRAQNIQSRLMGADT 116

QY 213 WFDSDCKFVLRTEFNKLAERGAET-----FKNIETMYTRIFLYSGEPTLYGNSTVFGPTG 268  
DB 117 RLD-----MSGFIEHKLTKNLE---EEIRARQGRKYI-----PAGASD 154

QY 269 NKTGLAIKRFYPPKPHLPTEKFLSLQIFDAVIV 305  
DB 155 HPLGGLGFAR--WAFVEVMEKE-----LGVFEDVIV 185

RESULT 8  
US-10-369-493-3172

Sequence 3172, Application US/10369493

Publication No. US2003023675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3172

LENGTH: 290

TYPE: PRT

ORGANISM: Neurospora crassa

US-10-369-493-3172

Query Match 5.0%; Score 96; DB 15; Length 290;  
Best Local Similarity 22.8%; Pred. No. 0.82;  
Matches 69; Conservative 40; Mismatches 94; Indels 100; Gaps 19;

QY 5 AGAARGASWCMALALMLAVVPGMSRVSGIPSR-----RHMPPYKRF--DFRKPDPYC 58  
DB 6 AGAALALASGVSAEAYI-----GFNSGNTLPSREAKPKQDWGEFTTAQTLKNSPGVN 59

QY 59 QAK-YT-----FCPTGSPIPVWE-----GDDI--EYFRLOAPWEEFYK 95  
DB 60 AVRLYTNTICAVSQRSEIEAFEALETNTKILGWMASGNTTIEPEIKALONGI--AKYG 117

QY 96 DLGLHKLIMDAIG-----FSTLTG-----KNYTWETELPQ--LGNCTF 134  
DB 118 KKLITL--IIGASIGSEDLVRSVYTGIONKSGVGAPALVYFIADWKKAFQGTAIANVPI 176

QY 135 PHL-----RPMDAFPWNGOACFPBSIDVHMKENGNTLYOVATISGNMFNOM 183  
DB 177 GHVDYTDAMTNGNKNEPVIDAVDW-----GVDDYFYEEKNGKNIN--SGYLDRA 226

QY 184 AK-----WVKQDNETGYETWNNVAKASPEKGAETWFDSDCKFVLRTEFNK 230

```
227 YDAIEGAVGKPIWV--TETGMPYVQGTWDQAAATINQOYYQWGVCRKL----FGKV 275
231 AEF 233
280 PTF 282

;SULT 9
-10-408-765A-58
Sequence 58, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Boia D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58
LENGTH: 836
TYPE: PRT
ORGANISM: Homo sapiens
-10-408-765A-58

Query Match 5.0%; Score 95.5; DB 16; Length 836;
Best Local Similarity 20.5%; Pred. No.3.8; Indels 73; Gaps 12;
Matches 61; Conservative 39; Mismatches 125;

67 TGSDIPV-----MEGDDIEVEFRLQAPVWEFKYGDLLG---HLKIMDAIGFRSTLTG 116
272 TSAASPAGVGTGMDDEDDPS-----KNFYSPSSPDKLTITFKTLRARESDDE 324
117 KNYTMEWEYLFQNGCTFPHLRPEMDAPFCNQGACFEGIDVHKENG-TLVQVATT 175
325 TQIKNMWEBAAGLLT--SLKDNV-----KATGVLYDVNMKYHMTGLTLREVSCK 376
176 SGNMENQAKWVKODNETGIYETWNVKASPEKGAETFDSDSKFLRTFNKLAERGA 235
377 LRRLQDHAENVQ---GALREIDDDIERFQKGSSTGYQ----- 415
236 EFKNIETN-YTRIFLYSGEPTYLGNETSVFGPTGNTKTLGLAIK-----RF 279
416 EMKQKAKULYQELLTQEGQASFGQKONVFGULRVVTFEFMKVKHLIDSLIDFLNPRF 475
280 YYPEKPHLPTEFLLSLQIFDAVIVHKQFY-----LFVNEFWELPMKPEFK 328
476 QPFGKPGIYTRERELTWF-IREVGVILSOVSKVHNGSEILFSYFODLVITLPELRK 532

;SULT 10
-10-282-122A-58749
Sequence 58749, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITTRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58749
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; -10-282-122A-58749

Query Match 4.9%; Score 94; DB 12; Length 1238;
Best Local Similarity 23.6%; Pred. No. 9.2;
Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

QY 110 FRSLTGNKTYMEWEYLFQNGCTFPHLRPEMDAPFCN-----QGACFEGIDVHK 164
DB 513 FTQYVGKNSALV-----NATTP-----WANGSIPKSNSTVRGGYGVWNG 555
QY 165 ENGLVQVAT-----ISGMENQAKWVKODNETGIYETWNVKASPEKGAETFDSDYC 219
DB 556 KTVGITGTFPADRYVITGNM-----TG-----NGAQTGGA--T 588
QY 220 SKYVLRTPFNKLAERGAFFKNIET--NYTRIFLYSGEPTYLGNETSVFGPTGNTKTLGLAI 276
DB 589 LNFVGAETEINIA--GATFKMLKITSONSYTFMALGD-----SSGSAKINVSQ 634
QY 277 KRFY-----YPEKPHLPTEFLLSLQIFDAVIVHKQFYLF-----YNFEYW-FLPMK 323
DB 635 SDFYDMTGGGDTDFGN-----GVFDSVFNKAYVYFQGTGENSEYNKKNINFLAGN 683
QY 324 FPFY-KITVEIPLPIRNKTLG 345
DB 684 FKFQKTTIEKSVLSDSASTPFDG 706

RESULT 11
-10-282-122A-55615
Sequence 55615, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
```



```

Y 154 FPEGIDVHKNGLTVQVATISGMNENOM--AKNV-----KONE 192
b 160 LYGDIDQMFVVDYITLTITKISICIFGKIVPAKXIMKTDDAFVRIDEVISLKSNS 219
Y 193 TGIYETWVWVWASP--EKGA-----ETW-FDSY-----DOSKFLRFENNL 230
b 220 NGLLYGLISQSSPHDKSKMFIISKKEPFEDYPPMAHGPCVVISRDIKAFVQCHOEL 279
Y 231 AERGAFFKNIETNYTRIIFYSGEPTVIGNETSVFGPTGNKTLGLAIKRFY 280
b 280 T-----LQFKLEDVAMGLMIQYKNSGGQVNIYVDKRFY 314

```

## RESULT 14

3-09-374-046A-160

Sequence 160, Application US/09374046A

Publication No. US20030096951A1

GENERAL INFORMATION:

```

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steindinger II, Robert J.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: GI 6075-83A
CURRENT APPLICATION NUMBER: US/09/374,046A
CURRENT FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 240
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 160
LENGTH: 742
TYPE: PRT
ORGANISM: Homo sapiens
-09-374-046A-160

```

```

Query Match 4.7%; Score 91; DB 10; Length 742;
Best Local Similarity 22.1%; Pred. No. 9.1;
Matches 62; Conservative 39; Mismatches 105; Indels 74; Gaps 16;

```

```

116 GKNTMWEYELFQGNCTFPH-----LRPMADP-FWCNQG---AACFEGIDVHWK 164
325 GSAVQFHSWRVFLV-VCAPFSVFAIGALTQPSBPREFLENGKHDAMWVLKQVHDTNMR 383
165 ENGLVQVATISG-NMFMOKAWKQDNENGIYETWVWVWASP-KGATWFDSDYCSKRFV 223
384 AKGHPERVSVTHIKTHQDEDELIEIOSDTGTWQRWGVRL-SLGGQVWGNFLSC----- 438
224 LRTEFNKLAEGAEKNIETN-----YTRIFLYSGEPTYL-----GNETSVP-- 264
439 -----FGPEYRRLITLMMGVMFTMSFSYGLTWFPDMIRHLQAVDYASRTKVPFG 489
265 -----GPTGNKTLGLAIKRFYPPKP-----HLPTKEFLSLIQ--IFDAVIHKOQFY--- 310
490 ERVEHVTNFETLENQIHRRGQYFNDKFIGRLKSVSFEDSLFEECYFEDVTSNTFFRNC 549
311 -----LFYN---FEYWFLEPMKFPFIKITX-----EELPLPI 338
550 TFINTVFYNTDLEPKYKFNRSR--LINSTFLANKEGCPLDV 587

```

SUET 15  
-10-616-263-160  
Sequence 160, Application US/10616263  
Publication No. US20040038276A1

```

/ GENERAL INFORMATION:
/ APPLICANT: Jacobs, Kenneth
/ APPLICANT: McCoy, John M.
/ APPLICANT: Lavallee, Edward R.
/ APPLICANT: Collins-Racie, Lisa A.
/ APPLICANT: Evans, Cheryl
/ APPLICANT: Merberg, David
/ APPLICANT: Treacy, Maurice
/ APPLICANT: Agostino, Michael J.
/ APPLICANT: Steindinger II, Robert J.
/ APPLICANT: Spaulding, Vikki
/ APPLICANT: Wong, Gordon G.
/ APPLICANT: Clark, Hilary
/ APPLICANT: Fechtel, Kim
/ APPLICANT: Genetics Institute, Inc.
/ TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
/ FILE REFERENCE: 00766.000103.5
/ CURRENT APPLICATION NUMBER: US/10/616,263
/ CURRENT FILING DATE: 2003-07-08
/ NUMBER OF SEQ ID NOS: 240
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 160
/ LENGTH: 742
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-616-263-160

```

```

Query Match 4.7%; Score 91; DB 12; Length 742;
Best Local Similarity 22.1%; Pred. No. 9.1;
Matches 62; Conservative 39; Mismatches 105; Indels 74; Gaps 16;

```

```

QY 116 GKNTMWEYELFQGNCTFPH-----LRPMADP-FWCNQG---AACFEGIDVHWK 164
DB 325 GSAVQFHSWRVFLV-VCAPFSVFAIGALTQPSBPREFLENGKHDAMWVLKQVHDTNMR 383
QY 165 ENGLVQVATISG-NMFMOKAWKQDNENGIYETWVWVWASP-KGATWFDSDYCSKRFV 223
DB 384 AKGHPERVSVTHIKTHQDEDELIEIOSDTGTWQRWGVRL-SLGGQVWGNFLSC----- 438
QY 224 LRTEFNKLAEGAEKNIETN-----YTRIFLYSGEPTYL-----GNETSVP-- 264
DB 439 -----FGPEYRRLITLMMGVMFTMSFSYGLTWFPDMIRHLQAVDYASRTKVPFG 489
QY 265 -----GPTGNKTLGLAIKRFYPPKP-----HLPTKEFLSLIQ--IFDAVIHKOQFY--- 310
DB 490 ERVEHVTNFETLENQIHRRGQYFNDKFIGRLKSVSFEDSLFEECYFEDVTSNTFFRNC 549
QY 311 -----LFYN---FEYWFLEPMKFPFIKITX-----EELPLPI 338
DB 550 TFINTVFYNTDLEPKYKFNRSR--LINSTFLANKEGCPLDV 587

```

Search completed: June 8, 2004, 14:11:56  
Job time : 72.1222 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 13:59:38 ; Search time 24.9889 Seconds

(Without alignments)  
1331.883 Million cell updates/sec

Title: US-10-010-050A-2

Perfect score: 1927  
Sequence: 1 MRGAGAGARASGCMALAL.....IKITYEIRPLPINKTISGL 346

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase:

1: PIR78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.5	5.3	3036	2 T18995	hypothetical prote
2	95.5	5.0	359	2 A96816	F9K20.25 (imported
3	94	4.9	362	2 T33904	hypothetical prote
4	94	4.9	1238	2 A64596	hypothetical prote
5	93	4.8	561	2 T19604	hypothetical prote
6	93	4.8	1327	2 B90674	hypothetical prote
7	93	4.8	1349	2 B85524	AIDA-1 adhesin-lik
8	92	4.8	502	2 F70316	probable beta-bar
9	91	4.7	291	2 T21702	conserved hypotet
10	91	4.7	508	2 D85056	hypothetical prote
11	90	4.7	585	1 UCI486	probable aspartic
12	89	4.6	1353	2 T27404	neopululanase (EC
13	88.5	4.6	465	2 T51094	hypothetical prote
14	88.5	4.6	752	2 E72616	acid phosphatase (
15	88.5	4.6	1462	1 DJHJAC	hypothetical prote
16	88	4.6	411	2 D69459	DNA-directed DNA p
17	88	4.6	742	2 E27263	conserved hypotet
18	89	4.6	3194	2 D71917	synaptic vesicle p
19	87.5	4.5	672	1 A41043	toxin-like outer m
20	87.5	4.5	755	2 T41912	acetate-CoA ligase
21	87.5	4.5	1162	2 T37889	structural phospho
22	87	4.5	619	2 T27026	probable nuclear p
23	86.5	4.5	388	2 T49253	hypothetical prote
24	86	4.5	411	2 AF1352	purple acid phosph
25	86	4.5	1086	2 JC6079	probable phosphoe
26	85.5	4.4	5035	1 I46646	chitin synthase (E
27	84.5	4.4	464	1 B59200	ryanodine receptor
28	84.5	4.4	465	1 T51095	acid phosphatase (
29	84	4.4	247	2 H69493	phosphoesterase-re

30	83.5	4.3	441	2 E71425	hypothetical prote
31	83.5	4.3	640	1 A30452	utromodulin precurs
32	83.5	4.3	984	2 T44496	cellulose 1,4-beta
33	83	4.3	228	2 H69399	hypothetical prote
34	83	4.3	552	2 E72283	alpha-galactosidas
35	83	4.3	620	2 T27008	hypothetical prote
36	83	4.3	653	2 F70383	organic solvent to
37	83	4.3	1086	2 T33893	hypothetical prote
38	83	4.3	4563	1 LEHUB	apolipoprotein B-1
39	82.5	4.3	403	2 AF3413	fosmidomycin resis
40	82.5	4.3	509	2 S49349	cyprosin (EC 3.4.2
41	82.5	4.3	704	2 B83914	beta-xylosidase (1
42	82	4.3	386	2 T02640	hypothetical prote
43	82	4.3	456	2 AH0842	probable glycopor
44	82	4.3	477	1 A36709	alpha-amylase (EC
45	82	4.3	553	2 T03858	hypothetical prote

ALIGNMENTS

RESULT 1  
T18995  
hypothetical protein C06B8.7 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T18995  
R/Steward, C.  
submitted to the EMBL Data Library, November 1996  
A/Reference number: Z19057  
A/Accession: T18995  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-3036 <WIL>  
A/Cross-references: EMBL:Z81463; PIDN:CAB03852.1; GSPDB:GN00023; CESP:C06B8.7  
A/Experimental source: clone C06B8  
C/Genetics:  
A/Gene: CESP:C06B8.7  
A/Map position: 5  
A/Intons: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/1

Query Match	5.3%	Score 101.5	DB 2	Length 3036;
Best Local Similarity	22.2%	Pred. No. 9.2;	Mismatches 143;	Indels 165;
Matches	103;	Conservative	52;	Gaps 30;
QY	10	GRASVCMALALLMLAVVPGWSR---	VSQIPSRHWPVYKRFDRPRDPCQAKYTF-	64
DB	1451	GRLSNRWNNKIMLQKV-NFTRNSEAVWMLKSPQAVP-----	CTPLAEITYHFD	1500
QY	65	-CP-TGSPFIPMEGDDI---	EVFRLOAPVWEKYG-----	DLI-----
DB	1501	NCSTVDNSGPIIEHSHRDLYASANVFHW--	ILMSNTFANNSRSGIAVALPDYTLAKQTH	1558
QY	99	-----GHKIMHDAIGFRSITLTGKNYMEWELFOLANGCTRP-----	HURPE	140
DB	1559	SFULTNFRPERSNFKILLDGYVAFANISSNNFTLN-----	TAPKQFGWELR-G	1607
QY	141	MDAPFCNOGAACPFESIDVHM--KENGCT---	LYOVATISGNMENOAKVXODNENGI	195
DB	1608	MEKVLICERNRPFENWNG---HMMIKLDATSQYLRQIDVDSYOVYNYTER-NRRTINRGD	1662	
QY	196	YIEYWNVKAPEKCAEFWPSYDCSKFVLRTFNKLAEP-----	GAEFKNI-ETNYT-----	245
DB	1663	YVDMW-----FRSYALGVFSQKVEVHFNFPEMLIDFEIIVSGAKYSDVFEIWNVTNNMW	1717	
QY	246	-----RIFLYSG-----	EPTVIGNETSY-----	PEPTGN-----
DB	1718	GTGNEAVIAQRVDFPDMMNTFTPAEMTPEYVSNDLSINFWNPNWRDQQLANATYLAEGVH	1777	
QY	270	-----XTLGIATKRF-----	YYPFKPHLPTEKEFLI---	SLIQIPDAVYH---K
DB	1778	DLHGRVYEDKNLTLITERWYSPFYNYRPFPRPRITRDVITMPATILYIEQNVGVHWPNV	1837	

308 QEFYLFYNF---EYWFLEPKFPFIKIT---YE-EIPLPIRNK 341  
 1838 RIIIVGNIVAGGEYV-QPIREKPIVTEYMQGKELPTEYRK 1879

## RESULT 2

196816

19K20.25 [Imported] - Arabidopsis thaliana

Species: Arabidopsis thaliana (mouse-ear cress)

Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

Accession: A96816

Theologos, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Hensen, N.F.; Hughes, B.; Huizart, L.

ature 408, 816-820, 2000

Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 Izzo, M.; Rooney, T.; Rowley, D.; Sakano, H.;  
 Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 er, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 Reference number: A86141; MUID:21016719; PMID:11130712

Accession: A96816

Status: preliminary

Molecule type: DNA

Residues: 1-359 <STO>

Cross-references: GB:AE005173; NID:g3834323; PIDN:AAC81039.1; GSPDB:GN00141

Gene: F9K20.25

Map position: 1

Query Match 5.0%; Score 95.5; DB 2; Length 359;  
 Best local similarity 21.5%; Pred. No. 2;

Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;

62 YTFCEPTGSDIPVMEGGDDIEFRLQAPVMEFYKGLGLKIMHAIQFRSLTGLKNTYM 121  
 80 YRMQPTGCDIPRFNGDFLTRKKGKILF---VGSLSN-----N 116  
 122 EYELFQAGNCTFFPHLRPEMDAPFCMNOGACFF---EGIDVHWKENGTLVQVAT--- 174  
 117 YWVSL---SCMLHAAPVAKYTFQNLKGLSTETPEYGL-SVNLKNGFLVDVSDKTR 171  
 175 ---ISGNMFMQAKVKKQDNENGIYETMNVKASPEKGAETPDSYDCKFVLRFNKCA 231  
 172 GLITLKDSTISRNQWLGSDVALENTFFHW---SHTGAKTW-DYFGTGDKXIVKEMNME 226  
 227 AFKIALTTWMSKWDHNDIPSKTRVYQGVSPVHLNGE---WGRKGTCLGRTVFWGSPS 283  
 281 YPRFPH 286  
 284 YPRGN 289

## SUBT 3

3904

Poetical protein Y25C1A.7a - Caenorhabditis elegans

Species: Caenorhabditis elegans

Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

Accession: T33904

Kalicki, J.; Smith, A.; Gibson, A.

Submitted to the EMBL Data Library, February 1999

Description: The sequence of C. elegans cosmid Y25C1A.

Reference number: Z21437

Accession: T33904

Status: preliminary; translated from GB/EMBL/DBD

Molecule type: DNA

Residues: 1-362 <KAL>

Cross-references: EMBL:AF125459; PIDN:AA012836.1; GSPDB:GN00020; CESP:Y25C1A.7a

Experimental source: strain Bristol N2; clone Y25C1A

A:Gene: CESP:Y25C1A.7a  
 A:Map position: 2  
 A:Insertions: 7/1: 30/3; 79/3; 146/1; 231/1; 282/3; 323/3

Query Match 4.9%; Score 94; DB 2; Length 362;

Best local similarity 21.7%; Pred. No. 2.7;

Matches 48; Conservative 27; Mismatches 88; Indels 58; Gaps 9;

QY 103 IMDAIGFSTLTGKNYTMWEYLF-----QIGNCTFPHLR-----PEMD 142  
 48 VFDSOSTSTKRNKFNFESEFYQQFDFVETDQVTKRLNSVIFTHRYVIGDFLPIDLM 107  
 QY 143 APFCMNOGACFPEGIDVHWKENGTLVQVATISGNMFMQAKVKKQDNENGIYETMNV 202  
 108 GPRVSV-----TLVPAIGIFGN-----LAQFINDGAKGTGSDFRM 145  
 QY 203 KASPEKGAE---TWFDSDYCS---KEVLRTEFKLAEPGAEPFNIEFTNYTRIFYSGEPT 255  
 146 AKIPEVTEISIFRAQFPKKNCKIIDFSMKKPKOKINNLCESPFKKSD-----FCHONSFS 200  
 QY 256 YLGNENSVFGPFGNKTGLAIR---FYPRFPHLPTEKFL 293  
 201 FC-NSHGFLTNGNLKTGSFSEQKNMFYTSFITSASTIFL 240

## RESULT 4

A64596

hypothetical protein HP0609 - Helicobacter pylori (strain 26695)

Species: Helicobacter pylori

Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

Accession: A64596

R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weldman, J.K.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Katrk, P.D.; Smith, H.O.; Fraser, C.A.

Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

Reference number: A64520; MUID:97394467; PMID:9252185

Accession: A64596

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

Residues: 1-1238 <TOM>

Cross-references: GB:AE000575; GB:AE000511; NID:g2313730; PIDN:AA07677.1; PID:g2313733

Query Match 4.9%; Score 94; DB 2; Length 1238;

Best local similarity 23.6%; Pred. No. 13;

Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

QY 110 FRSTLTGKNYTMWEYLFQAGNCTFPHLRPEMDAPFCN-----QGAACFPEGIDVHWK 164  
 513 FTQYGGKNSALVP-----NATTP-----WANGSIPKSNSTVRFGGYEGVNWG 555  
 Db 165 ENGLTVQVAT---ISGNMFMQAKVKKQDNENGIYETMNVKASPEKGAETWFDSDYDC 219  
 556 KTGYYITGFTADRYVITGNM-----TG-----NGAQGGGA--T 588  
 QY 220 SKFYLRTEFNKLAERGAEFKNIE---NYTRIFYSGEPTVIGNETSVFGPTGNKTGLAI 276  
 589 LNFVGAETVIA--GAFFKQLKTTSONSYTFMALGD-----SSGSXKIVNSQ 634  
 QY 277 KRFY-----YPRKPHLPTEKFLSLQIPDAVIVHKQFLF-----YNEFYW-FLPMK 323  
 635 SDFYDWTGGYDFTGN-----GVFDSVNPKNKAYIKFQGTENSTYFKNTNPLAGN 683  
 Db 324 PPRP-KTYEIRPLPIRNKTLG 345  
 Db 684 FKQGGKTTTEKSVLSDASYTFDG 706

## RESULT 5

T19604

hypothetical protein C31C9.4 - Caenorhabditis elegans

Species: Caenorhabditis elegans

```

Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #next_change 15-Oct-1999
Accession: T19604
Percy, C.
Submitted to the EMBL Data Library, December 1996
Reference number: Z19150
Accession: T19604
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-561 <M1>
Cross-references: EMBL:Z83219; PIDN:CAB05696.1; GSPDB:GN00020; CESP:C31C9.4
Experimental source: clone C31C9
Genetics:
Gene: CESP:C31C9.4
Map position: 2
Introns: 61/3; 98/1; 169/2; 197/2; 212/3; 246/2; 472/1; 510/2

Query Match
Best Local Similarity 20.7%; Pred. No. 5.7;
Matches 46; Conservative 41; Mismatches 69; Indels 66; Gaps 14;

y 156 EGIDV---HWKNGTIVQVATISGNMFMQAKWVKQDNETGIYETWNVKASPEKGA 211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 6 DGMEDLFEUDQKCN---AKTAHI-----ESTKYFRTETIEHLHFEHSISVNVK----- 49

y 212 TWFDSDCKFVLTNTKLAERGAEPKN--IENY-----TRIFYSPPTLTGNSTV 263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 50 -WFSYVDANK---VKKVLDSNNKPSGIFEVKYNCPLEIAQV---DPSAGDTSDF 99

y 264 FPGTGNKTL---GLAIKR-FYYPFKPHLPYKEFLSLQIFDAVIVHKQFYLFPNFEYWF 313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 100 LQEQMSVLTLENNPIARTCIIVFEFRKNLPIFESEFKNFGVLDGDLG-----PPEFEFWF 154

y 320 LPMKEPPIK---ITYE-----RIPPIRNKKTLSGL 346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 155 ---YRFYKGDIEDINYSQETTHPTLFDLPIDHDKILKNL 192

RESULT 6
96674
ada-1 adhesin-like protein [imported] - Escherichia coli (strain O157:H7, substrain R1K
|Species: Escherichia coli
|Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
|Accession: B90674
|Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
|Saawara, N.; Yasunaga, T.; Kohata, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
|NA Res. 8, 11-22, 2001
|Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
|Reference number: A99629; MUID:21156231; PMID:11258796
|Accession: B90674
|Status: preliminary
|Molecule type: DNA
|Residues: 1-1337 <HAY>
|Cross-references: GB:BA000007; PIDN:BAH33785.1; PTD:G13359819; GSPDB:GN00154
|Experimental source: strain O157:H7, substrain R1MD 0509952
|Genetics:
|Gene: ECS0362

Query Match
Best Local Similarity 20.3%; Pred. No. 17;
Matches 59; Conservative 32; Mismatches 93; Indels 106; Gaps 15;

y 1 MRRGAGARCARASMCWALMLLAVVPGMRSVSGIPRRHMPVYVKFRDPKPPYCOA 6C
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 944 LAKQAGNSG-----NWITSGSDSPHQPEP-----DPMNPNEPPNP 9E2

y 61 KYTFCEPTGSPDIPWEGDDDIEVFLQAPVWEFKYGLLGHILKT-----YADAIG-- 1C9
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 983 EPNNPNEPPTGPDINDVNDL---RPEA-----GSYIANDIAANTFTTIRHERLANTY 1C32

y 110 FRSITLG--KATYM-----EWYE-----LFQAGNCFPHLRPEMDAPFW 146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 1033 YTFDVTGEQKQTTMMWHEGSHNKRKDSGQGLTKQSNRYVLDGG-----DVAQW 1C62

```

```

QY      147  CNGCAACFEBSIDVHMKENGTLYQ-----VATISGMNMNQNAKWVKQD-NETGIYYE 198
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1083  SONGSDRWVGVMAAGVNSDSKTISRTPYAKAVSNVSTGLYATWVADDSRNGAYLD 1142
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      199  TANNKASPEKAEFTWFD-----SYDCKSFVLART---FNKLAEF 233
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1143  SW-----AQYSWFDNTVKGDLDQESYSKSGFTTASLEAGYKHKLAEF 1184
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
E85524
Probable beta-barrel outer membrane protein Z0402 [imported] - Escherichia coli (strain C)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85524
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A95480; MUID:21074935; PMID:11206551
A:Accession: E85524
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <STO>
A:Cross-references: GB:AE005174; NID:g12513130; PIDN:AA654657.1; GSPDB:GN00145; UMGF:Z0402
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0402

Query Match      4.8%; Score 93; DB 2; Length 1349;
Best Local Similarity 20.3%; Pred. No. 1;
Matches 59; Conservative 32; Mismatches 93; Indels 106; Gaps 15;

QY      1  MRGGAGARGRASWCWALALTLAVVPGSRVSGIPSRHWPVPYKRFDPKPDYCOA 60
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      966  LARGGANSQ-----NWYLTSGSDSPDLQEP-----DPMRPDEPNP 1004
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61  KTFEGTSPPIPVWBGDDIEVERLQAPVWEKYGDLGLKLI-----MHDAIG--- 109
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1005  EENRPPTPTPGDLDLVNDL-----RPEA-----GSYIANLAANTMTFTRLHERLGNTY 1054
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      110  FRSLTITG-KNYTM-----EWE-----LFQIGNCTFHLRPEMDAPFW 146
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1055  YTDVNTGEOKQTTMMNRHGHGKWRDGSQDKTOSNRVYLDLG-----DVAQW 1104
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      147  CNOGACAFEEGIDVHMKENGTLYQ-----VATISGMNMOMAKWKOD-NETGIYYE 198
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1105  SONGSDRWVGVMAAGVNSDSKTISRTPYAKAVSNVSTGLYATWVADDSRNGAYLD 1164
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      199  TANNKASPEKAEFTWFD-----SYDCKSFVLART---FNKLAEF 233
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1165  SW-----AQYSWFDNTVKGDLDQESYSKSGFTTASLEAGYKHKLAEF 1206
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
F70316
Conserved hypothetical protein aq_175 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: F70316
R:Deckerl, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Oye
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: F70316
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-502 <ROF>
A:Cross-references: GB:AE000677; NID:g2982900; PIDM:AA006534.1; PID:g2982916; GB:AE000677
A:Experimental source: strain VFS
C:Genetics:
A:Gene: aq_175

```

C:Superfamily: probable 60k inner membrane protein; stage III sporulation protein homold  
F:293-486/Domain: stage III sporulation protein homology <SPOR>

Query Match 4.8%; Score 92; DB 2; Length 502;  
Best Local Similarity 21.3%; Pred. No. 6;  
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

```
QY PIPWEGDDIEFRLQAPWVEFKYDGLGHKIMDAIGFSTLTGKNYTWEMWELFOL 129
Db 118 PLEIFGNDLDO-KINFEVEYELKRGK-NSVELIHKEIKVKITLISKNGAI----- 166
QY 130 GNCTEPHLRPE-MDAPFWCNQGA---ACFEGIDVHKENGTLVOVATISGNMNMOMA 184
Db 167 -----HLSVEGLKRPFWVFVSSPPDDEAFYTHVGPV-LKINGEVRLDV----- 209
QY 165 KWKQDNENGIYYETWVWVAKSPKCAETWFDSDSKFVLTFTFNKLAFFGAFFKNIETNY 214
Db 210 -----DDKGINPEGNIEFGSESRYPFGKAKOVQKHIVK---VLDGKFFVSLST-- 238
QY 245 TRIFLYSGEPT-YLG-----NENSVFGPTGNKTLGLAIK---RFYYPFKPHLPTEKFL 233
Db 259 ---FLYDEKITYLAKQYARLRGLVDTLDMGTLKTIIVKPLFLFYIYEH--TGSWV 313
QY 294 LSLT-----QIFDAVIVHKOPLYFYNFPEYWFLEPKKPFITKYEBEIPLRNKTU 343
Db 314 LSLVLTFTVIRIFLEPLGYSVSMQKQF--LAPKMKIKQKTKXDDVKKQSEEM 367
```

## RESULT 9

T21702  
Protein: hypothetical protein F33E2.5 - *Caenorhabditis elegans*  
Species: *Caenorhabditis elegans*  
Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
Accession: T21702

Submitted to the EMBL Data Library, January 1997  
Reference number: Z19461

Accession: T21702

Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA

Residues: 1-291 <MTL>

Cross-references: EMBL:284574; PIDN:CA806543.1; GSPDB:GN00019; CESP:F33E2.5

Experimental source: clone F33E2

Genetics:

Gene: CESP:F33E2.5

Map position: 1

Intons: 252/1

Query Match 4.7%; Score 91; DB 2; Length 291;  
Best Local Similarity 17.9%; Pred. No. 3.7;  
Matches 57; Conservative 47; Mismatches 92; Indels 122; Gaps 13;

```
Y 57 YCOAKYTFPCD---TSSPIPWEGDDIEFRL-QAPWVEFKYDGLGHKIMDAIGFR 111
b 25 YPEARVPTLPKIKTADFOKLGSKDRIAFILISQPIADKCTETIKVILBQ---PQ 80
Y 112 STLTKNY-----TWEMWELFOLANCTPHLRPEMDAPFWCNQGA 151
b 81 DTDIPNNTAVSNALNEFKIGONCHDFOPIDWYKNIK-GYFTYKTKTEGVOQAVFNDENG 133
Y 152 ACFFEGIDVHKENGTLVOVATISGNMNMOMAKWK--QDMETGYYETWVWAKSPKNG 203
b 140 NQGLAGV-----CYVATPAG-----KWAQKIQTOETLLRFEF--QKAEF-- 175
Y 210 AETWFDSDSKFVLTFTFNKLAFFGAFFKNIETNYRIFLYSGEPTLTGKNYTWEMWELF 269
b 177 ---FPUNSKSRFPDPAFKKFAFYRG---RYCELEGORVLWVGKA----- 214
Y 270 KTLGLAIKRFYYPFKPHLPTEKFLSLDQIFPAVIVHKOPLYFY-----NF 315
b 215 ---DGOIVHNVHVVYCMOWTKGCAEAKDHF 243
Y 316 EYWPLPKKFPFKITYEE 333
```

Db 244 EYWMIAPEBPVQKSKKE 261

## RESULT 10

DB5056

probable aspartic proteinase (imported) - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
Accession: DB5056

R:anonymus, The European Union *Arabidopsis* Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: A85001; MUID:20083488; PMID:10617198

Accession: DB5056

Status: preliminary

Molecule type: DNA

Residues: 1-508 <STO>

Cross-references: GB:NC\_001268; NID:g7267203; PIDN:CA877914.1; GSPDB:GN00140

Gene: AT4g04460

Map position: 4

Superfamily: oryzasin, saposin repeat homology

Query Match 4.7%; Score 91; DB 2; Length 508;  
Best Local Similarity 23.8%; Pred. No. 7.5;  
Matches 36; Conservative 17; Mismatches 48; Indels 50; Gaps 7;

```
QY 57 YCOAKY-----TFPCPTSPIPWEG-----DDIEFRLQAPWVEF----- 92
Db 125 YFHSKYVASQSSVRRKNGKXASTIRYGTGALSIGFSNDVAVGDIVKEQFIETATSEPGI 184
QY 93 -----KYGDLGHUKIMDAIGFSTLTGKNYTWEMWELFOLANCTPHLRPEMDAPFWC 147
Db 185 TFLIAKFDGILG-----LGFKEISVSG-NSIPVWVNVWEXG-----LVKEPIEFFWL 229
QY 148 N-----QGAACPFEGIDVHKENGTLVOY 172
Db 230 NRPKXDPGEGLVFGVDPGRHFGKHTFPV 260
```

## RESULT 11

JCI486

neopullulanase (EC 3.2.1.135) - *Thermoactinomyces vulgaris*

N:Alternate names: alpha-amyglase II

C:Species: *Thermoactinomyces vulgaris*

C>Date: 31-Dec-1993 #sequence\_revision 18-Aug-1995 #text\_change 16-Jun-2000

Accession: JCI486

R:Tonozuoka, T.; Ohtsuka, M.; Mogi, S.; Sakai, H.; Ohra, T.; Sakano, Y.

Biochem. Biotechnol. Biochem. 57, 395-401, 1993

A:Title: A neopullulanase-type alpha-amyglase gene from *Thermoactinomyces vulgaris* R-47.

A:Reference number: JCI486; MUID:93222535; PMID:7763540

Accession: JCI486

Molecule type: DNA

Residues: 1-585 <TON>

Cross-references: GB:D13178; NID:g391625; PIDN:BA02473.1; PID:g398125

Function:

Description: hydrolysis of alpha-(1->4)-glucosidic linkages of pullulan to produce pan

A:Pathway: pullulan degradation

A:Note: also has alpha-amyglase activity

C:Superfamily: neopullulanase; alpha-amyglase core homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:293-424/Domain: alpha-amyglase core homology <AMY>

F:329,354,421/Active site: Glu, Glu, Asp #status predicted

Query Match 4.7%; Score 90; DB 1; Length 585;  
Best Local Similarity 21.5%; Pred. No. 11;  
Matches 53; Conservative 30; Mismatches 68; Indels 96; Gaps 15;

```
QY 101 IKIMDAI-----GFRSTLT--GKNYTWEM-----YELFOLANCTF 134
Db 234 IKIIDAVFNHAGDQFAFRDVLQKGEOSRYKDWFTIEDFPVSTSRNVEYFAVQVAM 293
```

Y 135 PHLR-----PMD-----APFMCNQAACFEFG-----IDVHKNGNTGVYV----A 173  
b 294 PKLRTREVEKXELVDFARFMEOG-----IDGRFLDVANVDAFMKRPRLVXSINPDA 349  
Y 174 TIGSNFMNQAKVYKODNETGI-----YYETMNVKAS-----PEKG 209  
b 350 LIYGEIWIHDASGMVMDQFDSVNMVYLFRESVIRFFAIGELIHARFDLFRAMLIPECA 409  
Y 210 AE-TW--FDSYDCKFPLRTFNKLAERGAEPKXIEYTYTI-----FLYSGEP-TYLGNE 260  
b 410 AGLIMLIDSHDHYERFVLTSCGNGEAKF-----FLAVLFQMTYIGTPLYGDE 457  
Y 261 TSVFGPT 267  
b 458 IGWAGAT 464

## RESULT 12

27404

Ypothetical protein Y75B8A.22 - Caenorhabditis elegans

Species: Caenorhabditis elegans

Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

Accession: T27404

Submitted to the EMBL Data Library, November 1998

Reference number: Z20361

Accession: T27404

Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA

Residues: 1-1353 &lt;WIL&gt;

Cross-references: EMBL:AL033514; NID:e1343251; PIDN:CAA22106.1; CESP:Y75B8A.22

Experimental source: clone Y75B8A

Genetics:

Gene: CESP:Y75B8A.22

Intons: 28/1; 79/2; 634/3; 996/1; 1134/1; 1296/3

Query Match 4.6%; Score 89; DB 2; Length 1353;

Best Local Similarity 19.1%; Pred. No. 38;

Matches 73; Conservative 51; Mismatches 124; Indels 134; Gaps 16;

31 RVSGIP-----SRHMPVYKR--PDRPKPPYCO-----AK 61  
607 RSRGPPAAVGLYHARALWPSFKKGLDFDSPEEDQLQLEGLKADKKYAKDK 666  
62 YTFCTGSPFVPMEDDDIEVRLQAPVMEKYGILLHLKIMDAIGFSTLTGKNYTM 121  
667 AESCTCEDDPAKYCYDMDATALQS-LWE-QSTDTLARI-LSHELPSEST-SPVWVQL 722  
122 EWEELFQNGCTFPHLRPEMDAPF-----WCNQAACF 154  
723 D-----ITPDVQOKFAMLAIQPALRARDIPAAVGLYHSRKLMPGDEAIFG 768  
155 FEGIDVHKNGNTLVQVATISGNN---FNQAKWVK-----ODNETGIYYETMNVKAS 205  
769 APGI-----GYEEIHELKALLEDLHEVAREMKVAERADDPDEBPAEYDSEOE 820  
206 PEKAEFTW-----FDSYDCKFVLRTEFKLAERGAEPKXIEYTYTIYLYSGEPTYL 257  
821 EEEVPAWKVEEIDFQFDSYVC-----KFSNVVVLKMYVFLIN---DFS 861  
258 GNEISVSGPVGKNTGLGAIKRFYVPEKPHLPTKFFLLSLQIFPAVYVH-----KQ 308  
862 KNSHEL-----NQLVLMRLHRIADLK-LPIKLYQVSLQVFSKVENHEFTHLSKDLRKS 914  
309 FYLFYNEFYWFLPMKPPPIKIT 330  
915 SRLYEIYQFGFHLKKFESKET 936

## SUIT 13

1094

Id phosphatase (EC 3.1.3.2) purple 1, precursor [similarity] - sweet potato  
Species: Ipomoea batatas (sweet potato)

C/Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 22-Oct-2001  
C/Accession: T51094  
R/Durmus, A.; Bicken, C.; Spener, F.; Krebs, B.  
Biochim. Biophys. Acta 1434, 202-209, 1999  
A/Title: Cloning and comparative protein modeling of two purple acid phosphatase isozymes  
A/Reference number: Z25292; MUID:20028255; PMID:10556574  
C/Accession: T51094  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-465 <DUR>  
A/Cross-references: EMBL:AL006224; PIDN:CAA06921.1  
C/Genetics:  
A/Gene: pap  
C/Superfamily: kidney bean purple acid phosphatase; phosphoesterase core homology  
C/Keywords: phosphoric monoester hydrolase

Query Match 4.6%; Score 88.5; DB 2; Length 465;  
Best Local Similarity 21.4%; Pred. No. 11;

Matches 51; Conservative 31; Mismatches 77; Indels 79; Gaps 13;

QY 161 VHKNGNTLVQVATISGNN-----FNQAKWVKQ-----DNBTGIYFT-----W 200  
Db 87 VWSNSQHKVVA--KGNIRTYTFNYTSGYIHCTIRNLHYNTKXYEVGIGNTFSFW 144  
QY 201 NVKASPEKGAETWFDSDCKFVLRTEFNKLAERGAEPKXIE--TWYTR-----IFYSG 252  
Db 145 -FTPEPEYGPD-----VPYTFGLIGDLGSGFSNSRNLTHYERNPIRGAVLFEVG 192  
QY 253 EPTTYGN-----ETSVFGP-----TGKNTGLA-----IKRF 279  
Db 193 DLSYADNYPNHNDVNRWDYGRFVERSTAYQFPIWTAGNHEIDFAEIGETPKFPFTKRY 252  
QY 280 YYPEKPHLPTKEFLLSLQIFDAVYHKOQ--YLFYNEFYWFLPMKPPPIKITYEIP 335  
Db 253 HVPIYASGSTEFTFYISIKRASAYITIVLSYSAYGYTPQYWLBEELP--KVNRTETP 308

## RESULT 14

E72616

Ypothetical protein APE1390 - Aeropyrum pernix (strain K1)

C/Species: Aeropyrum pernix

C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C/Accession: E72616

R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hakiawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, C.; K.

A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum

A/Reference number: A72450; MUID:9310339; PMID:10382966

A/Accession: E72616

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-752 &lt;KAW&gt;

A/Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA480387.1; PID:01044173; PID:95105

A/Experimental source: strain K1

C/Genetics:  
A/Gene: APE1390

Query Match 4.6%; Score 88.5; DB 2; Length 752;  
Best Local Similarity 19.7%; Pred. No. 20;

Matches 67; Conservative 34; Mismatches 106; Indels 133; Gaps 18;

QY 20 LLMV-----AVPQMSRVSGIPSRHMPVYKRPDPKPPYCOAKYTFCTGSPFVME 75  
Db 180 LLMVGVYVYVYVWVVKGL--YHPTLDIRDLIS--DE----- 216  
QY 76 GDDIEVFRLOAPVMEKYGILLHLKIMDAIGFSTLTGKN-----YTWMEYELFOIGN 131  
Db 217 -----NAVVD-----LGFSGYSNGPWTWVDYFGWEDTD-- 247  
QY 132 CTFPHLRPEMDAP-----FMCNQAACFEFGIDVHKNGNTLVQVATISGNNFNQMA-K 185  
Db 248 -----PAADYETFGMTWHNIGQTLV-----TFWKDETT-----ELTPRLAVA 285



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 14.4167 Seconds

(without alignments)  
1249.684 Million cell updates/sec

file: US-10-010-050a-2  
object score: 1927  
sequence: 1 MRGAGARGRASWCMALAL.....IKITTEPIPLIRKMTISGL 346

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

imum DB seq length: 0  
imum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase : SwissProt\_42:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1927	100.0	407	1	CLN5_HUMAN
2	92	4.8	502	1	OXNA_AQUAE
3	91.5	4.7	1705	1	PRPV_MOUSE
4	90	4.7	585	1	NER2_THEVU
5	89.5	4.6	427	1	ARP3_ACACA
6	88.5	4.6	1462	1	DPOA_HUMAN
7	88	4.6	411	1	TRUD_ARCFU
8	87.5	4.5	672	1	ACSA_METSO
9	87.5	4.5	755	1	PI00_HSVTJ
10	87	4.5	1184	1	CHSD_EMENT
11	85.5	4.4	5035	1	RXR1_PIG
12	85	4.4	254	1	SSA2_MOUSE
13	83.5	4.3	352	1	A85C_MYCAV
14	83.5	4.3	564	1	TM16_HUMAN
15	83.5	4.3	640	1	URDM_HUMAN
16	83.5	4.3	2167	1	BPH1_YEAST
17	83	4.3	228	1	YC01_ARCFU
18	83	4.3	889	1	CI22_HUMAN
19	83	4.3	999	1	HGPB_HAEIN
20	83	4.3	1028	1	FDXG_HAEIN
21	83	4.3	1310	1	ACE_RABIT
22	83	4.3	4563	1	APB_HUMAN
23	82	4.3	2492	1	ATRX_HUMAN
24	82	4.3	2492	1	ATRX_PANTR
25	81.5	4.2	514	1	CHH2_SCHPO
26	81.5	4.2	606	1	VE1_HPV8
27	81.5	4.2	5037	1	RYR1_RABIT
28	81	4.2	644	1	URDM_RAT
29	81	4.2	1451	1	DPOA_RAT
30	80.5	4.2	282	1	HCHA_ECO57
31	80.5	4.2	544	1	OXNA_BORBU
32	80.5	4.2	555	1	POLG_DENIT
33	80.5	4.2	738	1	SEC6_DROME

34	80.5	4.2	835	1	VP3_ROTSL
35	80	4.2	429	1	CLN4_PHAVU
36	80	4.2	509	1	APR1_ORYSA
37	80	4.2	2492	1	ATRX_PONPY
38	79.5	4.1	511	1	SYFA_MERTH
39	79.5	4.1	603	1	VE1_HPV8
40	79	4.1	508	1	ASPR_HORVU
41	79	4.1	682	1	TDR5_HUMAN
42	79	4.1	928	1	ODO1_RICCN
43	79	4.1	2292	1	POLG_EMCVB
44	79	4.1	2292	1	POLG_EMCVD
45	78.5	4.1	218	1	VLYS_BPX3

## ALIGNMENTS

```

RESULT 1
CLN5_HUMAN
ID CLN5_HUMAN STANDARD; PRT; 407 AA.
AC 075503;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).
DN CLN5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., VARIANT VINCL ASN-279, AND VARIANT ARG-368.
RC TISSUE=Fetal brain;
RX MEDLINE=96324783; PubMed=9662406;
RA Savukoski M., Klockars T., Holmberg V., Santavuori P., Landier E.S.,
RA Peltonen L.;
RT "CLN5, a novel gene encoding a putative transmembrane protein mutated
RT in Finnish variant late infantile neuronal ceroid lipofuscinosis.";
RL Nat. Genet. 19:286-288(1998).
[2]
RN SUBCELLULAR LOCATION, AND GLYCOSYLATION.
RP MEDLINE=21968572; PubMed=11971870;
RX Isoomppi J., Vesa J., Jalanko A., Peltonen L.;
RT "Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5
RT protein.";
RL Hum. Mol. Genet. 11:885-891(2002).
CC -!- FUNCTION: Not known.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Glycosylated.
CC -!- DISEASE: Defects in CLN5 are the cause of Finnish variant late-
CC infantile neuronal ceroid lipofuscinosis (VINCL) [MIM:256731];
CC also known as ceroid lipofuscinosis neuronal 5 (CLN5). VINCL is a
CC fatal childhood neurodegenerative disease characterized by
CC progressive visual and mental decline, motor disturbance, epilepsy
CC and behavioral changes. The first symptom is motor clumsiness,
CC followed by progressive visual failure, mental and motor
CC deterioration and later by myoclonia and seizures.
CC -!- DATABASE: NAME=NCU CLN5;
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;
CC WWW="http://www.ucl.ac.uk/ncu/CLN5.html".
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
-----
DR EMBL; AF068227; AAC27614.1; -.
DR GenBank; HGNC:2076; CLN5.
DR MIM; 608102; -.

```



```

NM_: 256731; -
JR GO: GO:0016021; C: integral to membrane; TAS.
JR GO: GO:0008151; P: cell growth and/or maintenance; TAS.
CW Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
CW Disease mutation; Polymorphism; Epilepsy.
TRANSMEM 75
PT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT VARIANT 279 279 D -> N (in VLINCL).
/FTID=VAR 005137.
K -> R (in absNP:1800209).
/FTID=VAR 005138.
T VARIANT 368 368
T SEQUENCE 407 AA; 46339 MW; 449702D1DC9BFEE4 CRC64;
Q SEQUENCE 407 AA; 46339 MW; 449702D1DC9BFEE4 CRC64;

Query Match 100.0%; Score 1927; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 2, 8e-157;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MRRGAGAGRGASNCMALALMLAVDPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 6C
b 62 MRRGAGAGRGASNCMALALMLAVDPGMSRVSGIPSRHMPVYKRPDRPKDPYCOA 121
Y 61 KYTFCPTGSPIPVWESDDLEVPRLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNYT 120
b 122 KYTFCPTGSPIPVWESDDLEVPRLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNYT 181
Y 121 MEWELEFQNGCTFPHRPEMDAPFMCNOGAACFEFGIDDVHMKENGTLVQVATISGNMF 180
b 182 MEWELEFQNGCTFPHRPEMDAPFMCNOGAACFEFGIDDVHMKENGTLVQVATISGNMF 241
Y 181 NQAKWVKQONETGIYETNNVNASPEKGAETWFDSDCKFVLRTFNKLAEGAEFKNI 240
b 242 NQAKWVKQONETGIYETNNVNASPEKGAETWFDSDCKFVLRTFNKLAEGAEFKNI 301
Y 241 ETVYTRFLYSGBPTLYGNTSVGPFGNTLGLAIRFYPPKPHLPTKEFLSLQLT 300
b 302 ETVYTRFLYSGBPTLYGNTSVGPFGNTLGLAIRFYPPKPHLPTKEFLSLQLT 361
Y 301 DAVIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTLSGL 346
b 362 DAVIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTLSGL 407

RESULT 2
XAA AOUAE STANDARD; PRT; 502 AA.
D 06561;
T 15-MAR-2004 (Rel. 43, Created)
T 15-MAR-2004 (Rel. 43, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E Inner membrane protein oxaa.
N OXAA OR AQ 175.
S Aquifex aeolicus.
C Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
X NCBI_TaxId=63363;
P SEQUENCE FROM N.A.
C STRAIN=VF5;
X MEDLINE=98196666; PubMed=9537320;
A Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
A Graham D.E., Overbeek R., Shead M.A., Keller M., Aufay M., Huber R.,
A "The complete genome of the hyperthermophilic bacterium Aquifex
A aeolicus."
L Nature 392:353-358(1998).
I -I- FUNCTION: Required for the insertion of integral membrane proteins
I into the membrane. Probably plays an essential role in the

```

```

CC integration of proteins of the respiratory chain complexes.
CC Involved in integration of membrane proteins that insert
CC independently and independently of the Sec translocase complex (By
CC similarity).
CC -1- SUBUNIT: Specifically interacts with transmembrane segments of
CC nascent integral membrane proteins during membrane integration (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: Belongs to the OXA1/oxa family. Subfamily 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE00677; AAC06534.1; -.
DR PIR; P70316; F70316.
DR HAMAP; MF_01810; -.
DR InterPro; IPR001708; 60kDa innermem.
DR Pfam; PF02096; 60KD IMP; 1.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 12 34
FT TRANSMEM 308 330 POTENTIAL.
FT TRANSMEM 378 400 POTENTIAL.
FT TRANSMEM 447 469 POTENTIAL.
SQ SEQUENCE 502 AA; 58224 MW; 6C24B8405F2570AD CRC64;

Query Match 4.8%; Score 92; DB 1; Length 502;
Best Local Similarity 21.3%; Pred. No. 2, 6;
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

QY 70 PIPVMEGDDIEVFLQAPVWEFKYGDILGLKIMHDAIGRSTLTGKNYTWMEYLFOL 129
Db 118 PLEIFGNDLDDQ-KNPFEEYEIKEGK--NSVELIKHEKLVKILSYKCAI----- 166
QY 130 GNCTFPHARE-MDAFPWNOGA-----ACFEFGIDDVHMKENGTLVQVATISGNMFNOQA 184
Db 167 -----HLSVEGLKPEFWFVGSPPDDEAFYTHGVV-LKINGEAVRLDV----- 209
QY 185 KMWKQDNETGIYETNNVNASPEKGAETWFDSDCKFVLRTFNKLAEGAEFKNIETNY 244
Db 210 -----DOLKGINEFBGNIFFGGEBSRYFFKAKDYOKHIVYK---VKLGDKFVSLST-- 258
QY 245 TRIFLYSGEPT-YLG-----NETSVFGPTGNTLGLAIK--PRYYPFKPHLPTKEFL 293
Db 259 ---FLYDGEXTIYLGAQDYARLRELGLVDTLDWGTLKIVKFLFLVYIYEH--TGSMV 313
QY 294 LSLV-----QIFDAIVHKQFLFYNFEXWFLPMKFPFIKITYEELPIPIRKNTL 343
Db 314 LSLVLTFLVIRIFLPLPGYKSVSMQKQE--LAPRMEXIKOKYKDDPVKQOEEMW 367

RESULT 3
PTPV_MOUSE STANDARD; PRT; 1705 AA.
ID PTPV_MOUSE
AC P70289;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Receptor-type protein-tyrosine phosphatase V precursor (EC 3.1.3.48)
DE (Embryonic stem cell protein-tyrosine phosphatase) (ES cell
DE phosphatase).
GN PTPV OR ESP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RN SEQUENCE FROM N.A.

```

Query Match	Best Local Similarity	Score	DB 1;	Length	1705;
Matches	78; Conservative	21.4%;	38; Mismatches	136; Indels	113; Gaps
FT CARBOHYD	345	345	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	431	431	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	551	551	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	570	570	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	620	620	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	649	649	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	663	663	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	737	737	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	851	851	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	882	882	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	970	970	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
FT CARBOHYD	982	982	N-LINKED	(GLCNAC. .)	(POTENTIAL. .)
SQ SEQUENCE	1705 AA;	186795 MM;	2783755F15387D5B	CRC64;	
Query Match	4.7%;	Score 91.5;	DB 1;	Length 1705;	
Best Local Similarity	21.4%;	Pred. No. 13;			
Matches	78; Conservative	38; Mismatches	136; Indels	113; Gaps	1
QY 2	BRGAG-----AARGASCMWALALLMLAVVPGW-----SRVSGTIPSR-----	39			
DB 32	KGGGPPPLSVSVTSRGRP--SLFISWAAERGGFDVALCLRAMNLGPEEGQOLQAH	88			
QY 40	-----HMPVPEKFD-----FRPKPDYCOAKYFPC-----PTGSPI 71				
DB 89	NESSEFPEHGLVGRSRQLELVLRP-----CWQNVITITLARTAPTVYRGLQLHSTSPA	14			
QY 72	PV-----MEGDDLVFVFLQAP-----VMEFKGDLGLKIMDAIGFR 111				
DB 144	SLBASWSDASGGQSDYQLLLHPHESHTTLCNVSVSPDITLSPNGVLLPGSQYVLEVTWA	203			
QY 112	SLTNKNTMWEVLEFQLGNCFTFPHLRPEMDAPFW--CNOGAACFEFGIDVHKKENTLV	170			
DB 204	GLHAKTSLIQWTEVPVPHLAVRALGTSLSLQAFMNSSGATVFLITDLEGTULTKV	263			
QY 171	QVATISGNNENOMAKVAKODNETGIY-----ETWYKASPEKGAETWPFSDYC	219			
DB 264	VROGISTHTFRLS-----PGTPYQLKICAAAGPHQIWGNAT-----EWTYPSYPS	310			
QY 220	SKFVLRTEKNKLAIEFGAEFPKNLETNTRIFLYSG---EPIYLGNE--TSVFG--PGNKNT	271			
DB 311	DVLTFPLNMLMAMKAKAGGARDGY--VKLISGPVENTTTLGPBECNAVFPGLPGHYT	368			
QY 272	LGAI 276				
DB 369	LGRLV 373				
RESULT 4					
ID NEP2.THEVU	STANDARD;	PRT;	585 AA.		
AC Q08751;					
DT 01-NOV-1995 (Rel. 32, Created)					
DT 01-NOV-1995 (Rel. 32, Last sequence update)					
DT 15-MAR-2004 (Rel. 43, Last annotation update)					
DE Neopululanase 2 (EC 3.2.1.135) (Alpha-amylase II) (TVA II).					
OS TVAIL.					
ON Thermactinomyces vulgaris.					
OC Bacteria; Firmicutes; Bacillales; Thermactinomycetaceae;					
OC Thermactinomyces.					
OX NCBI_TaxID=2026;					
RN RN					
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.					
RC STRAIN=R-47;					
RX MEDLINE=93222535; PubMed=7763540;					
RA Tomozuka T., Ohtsuka M., Mogi S.-I., Sakai H., Ohta T., Sakano Y.;					
RT "A neopululanase-cyde alpha-amylase gene from Thermactinomyces					
RT vulgaris R-47."					
RL Biosci. Biotechnol. Biochem. 57:395-401(1993).					
RA Kamitani S., Kondo S., Okuyama K., Yokota T., Shimura Y., Tomozuka T.					
RA MEDLINE=99241045; PubMed=10222200;					
RA X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).					
RP STRAIN=R-47;					
RC MEDLINE=99241045; PubMed=10222200;					



```

1 STRAND 553 555
2 STRAND 561 561
3 STRAND 564 565
4 STRAND 566 567
5 STRAND 568 573
6 STRAND 575 576
7 STRAND 578 582
8 SEQUENCE 585 AA; 67467 MW; E311813A05A7791A CRC64;

Query Match
Best Local Similarity 21.5%; Score 90; DB 1; Length 585;
Matches 53; Conservative 30; Mismatches 68; Indels 96; Gaps 15;

101 IKIMEDAI-----GKSTLT--GKNYTWEM-----YELFOLGNCFT 134
234 IKIILDAVNHAGDGFPAFRDVLQKGQGRYKDWFIEDFPVSKSRNYEFNFVQVPM 293
135 PHLR--PMD-----AFQCNQACFPEG-----IDVHKENGTLVQV---A 173
234 PLKLRTEPEVKELPLDVAFRMEQG---IDGMRLDVANEVHAFMRFRRLVKSINPDA 349
174 TISGMNFNOMAKVVKQDNETGI-----YYETMNVKAS-----PEKG 209
350 LIVGEIWDHDSGLMGDOPDSVMNTLFRBSVIRFPATGEIHAERPDALITRAMLYPQQA 409
210 AE-TV--PDSDYCKFVLRTFNKLAFFGAFFKNIBNTYTRI-----FLYSGEP--TYLGNE 260
410 AAGLWNLDSHDTFRITNSCGNEAKF-----RLAVLPQMTYLTGLPLIYDDE 457
261 TSVFPGPT 267
458 IGMAGAT 464

RESULT 5
3P3 ACACA STANDARD; PRT; 427 AA.
ARP3 ACACA
P53490;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
15-JUL-1999 (Rel. 39, Last annotation update)
Actin-like protein 3.
ARP3.
Acanthamoeba castellanii (Amoeba).
Eukaryota; Acanthamoebidae; Acanthamoeba.
NCBI_TaxID=5755;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 30010 / Neff;
MEDLINE=96017709; PubMed=7593166;
Kelleher J.F., Atkinson S.J., Pollard T.D.;
"Sequences, structural models, and cellular localization of the
actin-related proteins Arp2 and Arp3 from Acanthamoeba.";
J. Cell Biol. 131:385-397(1995).
-!- FUNCTION: Part of a complex implicated in the control of actin
polymerization in cells (By similarity).
-!- SUBUNIT: Belongs to a complex composed of ARP2, ARP3, P41-ARC,
P34-ARC, P21-ARC, P20-ARC and P16-ARC (By similarity).
-!- SIMILARITY: Belongs to the actin family. ARP3 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U29610; AAA93068.1;
InterPro; IPR004000; Actin_like.
Pfam; PF00022; actin; 1.
SMART; SM00269; ACTIN; 1.
PROSITE; PS01132; ACTINS_ACT_LIKE; 1.

```

```

KW Structural protein; Cytoskeleton.
SQ SEQUENCE 427 AA; 48635 MW; A53A0965B1E7AP6 CRC64;

Query Match
Best Local Similarity 4.6%; Score 89.5; DB 1; Length 427;
Matches 64; Conservative 49; Mismatches 96; Indels 101; Gaps 17;

55 DRYCAKATFCP--TGSPFIPVME---GDDLE---VFRQLQAVMEPFKGDILGHKIMKD 106
44 DPCRRARWCMWPAAGKNIADIDFISDBAYNSKYVQITMFV---RRGV----- 91
107 AIGFRSTLTGKNY--MEWYELFOLGNCFTPHLR-----PEMDAPFMCQGAAC 153
92 -----ENWTHMEGF---WEHCIFKLRCPEDHFFLLTPEPLNAPNREYTAI 137
154 FPEGIDVHKENGTLVQVATISGMNFNOMAKVVKQDNETGIYETW----- 200
138 MFE-----TFNVPGIYIVQAVLALAAWTSQVYKELTGLTVISGQGVTHVIPABEGY 192
201 ---NVKASPEKGAE--TFE-----DSYCKFVLRTFNKLA--EFGARFKN 239
193 VIGSISKIIPLAGRDITTFVQLRERNEKIPPAETLEVAKAIKETFSVCPDIYKEFKK 252
240 IETNYTRIFLYSGEPTVIGNETSVFPGTGNK--TLGLAIKFFYVP--FKPHLPKPEPL 294
253 YTEPDKWF-----KTYEGIES-----VGKKPVNDVGVERLGEIPEIFSSDFLT 302
295 SLQIFDAVI 304
303 PLPKVDETI 312

RESULT 6
DPOA HUMAN
ID DPOA HUMAN STANDARD; PRT; 1462 AA.
AC P09884;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7).
GN POLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=88196090; PubMed=3359994;
Wong S.W., Wahl A.F., Yuan P.-M., Arai N., Pearson B.E., Arai K.,
Korn D., Hunkapiller M.W., Wang T.S.-F.;
"Human DNA polymerase alpha gene expression is cell proliferation
dependent and its primary structure is similar to both prokaryotic
and eukaryotic replicative DNA polymerases.";
EMBO J. 7:37-47(1988).
[2]
SEQUENCE OF 1-8 FROM N.A.
MEDLINE=91172197; PubMed=2005899;
Pearson B.E., Neshener H.P., Wang T.S.;
"Human DNA polymerase alpha gene: sequences controlling expression in
cycling and serum-stimulated cells.";
Mol. Cell. Biol. 11:2081-2095(1991).
-!- FUNCTION: Polymerase alpha in a complex with DNA primase is a
replicative polymerase.
-!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ [DNA] (N).
-!- SUBCELLULAR LOCATION: Nuclear.
-!- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
alpha, beta, gamma, delta, and epsilon which are responsible for
different reactions of DNA synthesis.
-!- SIMILARITY: Belongs to the DNA polymerase type-B family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X06745; CAA29920.1; --  
EMBL; M64481; AAA52318.1; --  
PIR; S00257; DJHUC.  
Genew; HGNC:9173; POLA.  
CK; p09884; --  
XIM; 312040; --  
GO; GO:0005634; C:nucleus; NAS.  
GO; GO:000389; F:alpha DNA polymerase activity; NAS.  
GO; GO:0006260; P:DNA replication; NAS.  
InterPro; IPR006172; DNA\_pol\_B.  
InterPro; IPR006134; DNA\_pol\_B\_dom.  
InterPro; IPR006133; DNA\_pol\_B\_exo.  
InterPro; IPR004578; Pol2.  
Pfam; PF00136; DNA\_pol\_B; 1.  
Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
PRINTS; PR00106; DNAPOLB.  
SMART; SMO0486; POLBc; 1.  
TIGRPFAMs; TIGR00592; pol2; 1.  
PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
Transferase; DNA-directed DNA polymerase; DNA replication;  
DNA-binding; Nuclear protein.  
DNA\_BIND 650 715 POTENTIAL.  
DNA\_BIND 1245 1376  
SEQUENCE 1462 AA; 165860 MW; 25C270B0A0DB38B8 CRC64;

Query Match Best Local Similarity 4.6%; Score 88.5; DB 1; Length 1462;  
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

46 KRFDFRKPDPYCAKTFPPTG-----SPFVMEGD--DIEV---FRL 85  
183 KKRISGASPPFVHTATVAPSGKIASPVSRKPPPLPVPLKRAFGADVQESTEEQ 242  
86 QAVWEKKYVDLGHKIMDAIGFSTLGNKNTVMWYELFQNGTFFHLAPENDAP 145  
243 ESGAMEEDDDP-----DEPVEVEV-----DLRP-MAAKA 272  
146 WCNQGAACFPEGIDVWKEN---GLVQVATISGMFNOMAKM-VKQDNETIYYETWN 201  
273 WDKES-----EPAEVYQGEADSGKTV---SYLGSFLPVSQMDIDQEDSSFSVQEVQ 323  
202 VKASP---EKGAT-----WFDSD--CSKFVLRKFNKLABFGAE-----FKNI 240  
324 VDSSTHPLVAGADEQVHFHYMDAYEDQYNOGVVFLFGKVMIESAEHVSQCVWVKNI 383  
241 ETMYTRFLVSGEPTVGNETSVEGPTGNT-LGLAKRFYPPKPHLPKKEFLSLQL 239  
384 ERT-----LY-----FIPRMKIDLTNGKGTGPIISKDYEEFDEKATKYIKM--- 429  
300 FDAVIVKQFYLFNFEYWFPMKFFPKITYE---EILPPIRNKTL 344  
430 FKSMPVKM---YAFELIPVPESELEVKYSAEMVQLDQDKGERFS 474

RESULT 7  
TRUD\_ARCFU STANDARD; PRT; 411 AA.

10-OCT-2003 (Rel. 42, Created)  
10-OCT-2003 (Rel. 42, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Probable tRNA pseudouridine synthase D (EC 4.2.1.70) (Pseudouridylylate synthase) (Uracil hydrolyase).  
TRUD OR APL677.

Archaeoglobus fulgidus.  
Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
Archaeoglobaceae; Archaeoglobus.

NCBI TaxID=2234;  
[1] --  
SEQUENCE FROM N.A.  
SPRAIN-VC-16 / DSM 4304 / ATCC 49558;  
MEDLINE=98049343; PubMed=9389475;

Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
Rieschmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodex A., Zhou L.,  
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
Colton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,  
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
Venter J.C.;  
"The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";  
Nature 390:364-370(1997).  
-1- FUNCTION: Could be responsible for synthesis of pseudouridine from uracil-13 in transfer RNAs (By similarity).  
-1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine 5'-phosphate + H(2)O.  
-1- SIMILARITY: Belongs to the pseudouridine synthase trnd family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AB000987; AAB89568.1; --  
PIR; D69459; D69459.  
TIGR; APL677; --  
HAMAP; MF\_01082; --; 1.  
InterPro; IPR001656; UPF0024.  
Pfam; PF01142; UPF0024; 1.  
TIGRPFAMs; TIGR00094; TIGR00094; 1.  
PROSITE; PS01268; UPF0024; 1.  
DR tRNA processing; Lyase; Complete proteome.  
KW ACT SITE 81  
ACT SITE 81 BY SIMILARITY.  
SEQUENCE 411 AA; 47745 MW; 109B5250D969922E CRC64;

Query Match Best Local Similarity 4.6%; Score 88; DB 1; Length 411;  
Matches 55; Conservative 45; Mismatches 97; Indels 74; Gaps 13;

74 MEGDDIEVFRLQAPVWEKYGDLGHKIMDAIG-----FRSTLTGKNYTMWYEL 126  
38 LSGDEGFLIIRVEKKM-----DTLNFARVLNSALGISQKRISFAGTKOKRALTYGYFSI 92  
127 FQNGTCTFHLAPENDAPWCNQGAACFPEGIDVWKE-----NGLVQVATISGN 178  
93 YGVKK-----EELERVLMKQAKIEVTIGYARRAQLDGLLN 128  
179 MEMQAKWKQDNENETIYYETWNVKSAPKGAETWF--DSYDCKFLVLTPTNGLAEGAE 236  
129 FF-RIRVYGRDGE--IFQETRN--ELMEKGTNFGGLRFGSIRITTEVGKLI----- 178  
237 FKNIEETNYTRIF-LVSGEPTVIGNE-----TSVFQPTGNTKTLGLAKRFYPPKPHLPK 290  
179 ---LQNNYBEAFVYVAKFEGENEVRKIRIELMETROAKIGRLFPYLRVRYRNLQK 235  
291 -----EFLSLLOIFDAVYHK-QFYL 312  
236 LREGKSEEBALSLPRLKLMFVHAQSYIF 266

RESULT 8  
ACSA\_METSO STANDARD; PRT; 672 AA.  
ID ACSA\_METSO

```

Db      341 VVAPELIGTSLMBSGAADYPPGKWKXKIQDHKATVLTATATAYRMFKQGAE--MPDKY 399
QY      218 DQSKFYLRPFNKLAE-----FG-AEFONIET--NYTRIPLYS-----251
Db      400 DLS--LRLLGSVGEHPINDEAMWYREHIGSELQIMDTWQTEGTFLNSPLPITPLPK 457
QY      252 GEPTY--LGENSVSGPNCKNT-LG-----LAIKFFYPFKHLEPKTEPLSLQITF--D 301
Db      458 GGCETFLPFDYDLSIDEBENEPDLSGGNVALKP--YP-----SMLRAFWD 503
QY      302 AVIVHKQFYLFFYNFEYFLP 321
Db      504 KERFMKEWQF-----YMDVP 519

RESULT 9
ID      P100 HSVJ7J STANDARD; PRT; 755 AA.
AC      P52519;
DT      01-OCT-1996 (Rel. 34 Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Large structural phosphoprotein homolog (Pp100).
GN      U11.
OS      Human herpesvirus (type 7 / strain U1) (HHV7).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Simplexvirus.
OX      NCBI_Taxid=57278;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Nicholas J.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC      -! SIMILARITY: TO THE LARGE STRUCTURAL PHOSPHOPROTEINS OF HSV-6 AND
CC      HCMV UL32.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed, usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U43400; AAC54672.1; -.
DR      PIR; T41912; T41912.
KW      Matrix protein; Phosphorylation.
SQ      SEQUENCE 755 AA; 86580 MW; 4083744CCCF91DA CRC64;

Query Match 4.5%; Score 87.5; DB 1; Length 755;
Best local Similarity 20.7%; Pred. No. 10;
Matches 58; Conservative 37; Mismatches 98; Indels 87; Gaps 15;

QY      144 PF-WCNGAGAC---PREGIDD--VHKKN-GTIVQVATISGN-----MNQNAKVV 187
Db      7 PFAVITSEACQELSRFPENISSLPVVDIRENPWILSQCIWKTGNSINNVKTLVNNILW- 65
QY      188 KODNETIYYEFWNVVKASPEKGAETWEDSYDCKFV-----LRTFNKLAE 232
Db      66 -----LYHQQLCKKKPDY-EEWQCEILKQKILKDYLQEQOMITDYSLSLSPKVG- 116
QY      233 FGAEFKQIEINNYRL--FLYSGEPITYGN-----ETSVFG--PTGNKTGLAIKRF 279
Db      117 FEEFEKQVARDLKLGSFLRMGVTHAADVNLITTBRAIGNLQAKANNMISFTIYQI 176
QY      280 YPEFKPF-----LPTKSELSILO-----IPDAIVHKQFY 310
Db      177 VDPMNENGVVYVNNLLYIGNLLITLHSGMMMEKALNTIEKKALIKALENNKVF 236
QY      311 LFFNFEYFLPMK---FPPIKITYEIDLPINNKTLSGI 346
Db      237 SIYSQYQLSLPLTSHRWTSFFKILTEDPQVITKSEIHAL 276

```

```

RESULT 10
HSD_EMENTI STANDARD; PRT; 1184 AA.
D CHSD_EMENTI
C P78611; 000744;
T 15-JUL-1999 (Rel. 38, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Chitin synthase D (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
transferase D) (Class-V chitin synthase D).
N CHSD OR CHSE.
E Emericella nidulans (Aspergillus nidulans).
S Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
C Eurotiales; Trichocomaceae; Emericella.
X NCBI_TaxID=162425;
N N
N SEQUENCE FROM N.A.
P STRAIN=FGSC 89;
C MEDLINE=96285566; PubMed=8709948;
A Mochizuma T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
"the Aspergillus nidulans genes chsd and chsd encode chitin synthases
which have redundant functions in conidia formation.";
Mol. Gen. Genet. 251:442-450(1996).
[2]
ERRATUM.
P MEDLINE=97188530; PubMed=9037115;
A Mochizuma T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
Mol. Gen. Genet. 253:520-528(1997).
[3]
REVISIONS.
P Mochizuma T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
P STRAIN=FGSC 4;
C MEDLINE=96406387; PubMed=8810520;
A Specht C.A., Liu Y., Robbins P.W., Bulawa C.E., Iartchouk N.,
Winter K.R., Riggall P.C., Rhodes J.C., Dodge C.L., Culp D.W.,
Borgia P.T.;
"The chsd and chse genes of Aspergillus nidulans and their roles in
chitin synthesis.";
Fungal Genet. Biol. 20:153-167(1996).
T Fungal Genet. Biol. 20:153-167(1996).
C - FUNCTION: Plays a major role in cell wall biogenesis.
- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + [(1,4)-(N-acetyl-
glucosaminyl)](N) = UDP + [(1,4)-(N-acetyl-D-
glucosaminyl)](N+1).
C - SUBCELLULAR LOCATION: Plasma membrane-bound.
C - SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUPERFAMILY CLASS
V.
-----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
C EMBL; D83246; BA011866.2; -
C EMBL; U52362; AAA97482.1; -
C InterPro; IPR004835; Chitin_synth_fng.
C InterPro; IPR001199; Cyt_B5.
C Pfam; PF03142; Chitin_synth_2; 1.
C Pfam; PF00173; heme_1; 1.
C Transferrase; Glycosyltransferase; Transmembrane; Cell wall;
M Multigene family.
T TRANSMEM 221 241
T TRANSMEM 476 486
T TRANSMEM 1039 1059
T TRANSMEM 1073 1093
T TRANSMEM 1097 1117
T TRANSMEM 1184 AA; 133504 MM; D6157184F154ED55 CRC64;
Q SEQUENCE

```

```

Query Match 4.5%; Score 87; DB 1; Length 1184;
Best Local Similarity 20.5%; Pred. No. 20;
Matches 75; Conservative 44; Mismatches 132; Indels 114; Gaps 20;

QY 63 TRCPETSPLEPWEGDDIVF---RLQAPVBEFKYKDLGHKIMHDAIGERS---TLT 115
D 189 TYC---SVITFMAPPFLVLCFCGMPKACRSANREKIG-LISILIMIAFVGFLTFGTAT 244
QY 116 -----GKNYMEWELEFQNGCTEPHNR--PEM-----DAPF-WONGAAC 153
D 245 VCGTPTRLKINIGSGVWIFHGQAYDLTKSTHPAAAGIPDMTNVLYDIPHKYGGDGSF 304
QY 154 FREGID-----DVHWKENGTLVQVATISGNMFQMAKVKQDNETGIYETW 200
D 305 PFGEVVGAGKGLITRENDSDIPTNSGDLAW--PPCHAFNDGS--SEPTVTSTYNGW 360
QY 201 --NVKASPEKGAE-----TFDSYDCKFVLRITNKLAEGAE-----FKN 239
D 361 ACHTSGSARKSFYSLKNSGDVYFTWEDTYNTR--KLAVYSGNVLDLNLNWFDD 413
QY 240 IETNYTRIF-----LYSGEPTYLG--NETSVFEGPTGNKTLG-LAIR 278
D 414 TQVNYPTKRLRDNDIDINGVDLTYFTQGEDKQIKCSQIITKVGSIIDTDTVGCIASQV 473
QY 279 FYYPFKPHLPTEKFLLSLQIPDAVIVHKQFYLFYFEXWFLPMKFPFKITYERIPLP 338
D 474 VLY-----VSLIFITISIVVKPAFALLPQ--WFLAPRPAQKTSWGAVDSKA 518
QY 339 RNKTL 343
D 519 RNOQT 523

RESULT 11
RYR1_PIG STANDARD; PRT; 5035 AA.
AC P16960;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR CRC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Norwegian Landrace; TISSUE=Skeletal muscle;
RX MEDLINE=93036581; PubMed=1329581;
RA Harbitz I., Kristensen T., Bosnes M., Kran S., Davies W.;
RT "DNA sequence of the skeletal muscle calcium release channel cDNA and
RT verification of the Arg615-->Cys615 mutation, associated with porcine
RT malignant hyperthermia, in Norwegian Landrace pigs.";
RL Anim. Genet. 23:395-402(1992).
RN [2]
RP SEQUENCE OF 1129-2801 FROM N.A.
RA Brenig B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1129-2643 FROM N.A.
RC STRAIN=German Landrace; TISSUE=Liver;
RX MEDLINE=94117003; PubMed=8288238;
RA Leeb T., Schmolzi S., Brem G., Brenig B.;
RT "Genomic organization of the porcine skeletal muscle ryanodine
RT receptor (RYR1) gene coding region 4624 to 7923.";
RL Genomics 18:349-354(1993).
RN [4]
RP SEQUENCE OF 4785-5035 FROM N.A.
RX MEDLINE=9105640; PubMed=2174405;
RA Harbitz I., Chowdhary B., Thomsen P.D., Davies W., Kaufman U.,
RA Kran S., Gustavsson I., Christensen K., Hauge J.G.;

```



"Assignment of the porcine calcium release channel gene, a candidate for the malignant hyperthermia locus, to the 6p11-->q21 segment of chromosome 6."

- GENOMICS 8:243-248 (1990).
- FUNCTION: Communication between transverse-tubules and sarcoplasmic reticulum. Contraction of skeletal muscle is triggered by release of calcium ions from SR following depolarization of T-tubules.
- SUBUNIT: Homotrimer (Potential).
- MISCELLANEOUS: The calcium release channel is modulated by calcium ions, magnesium ions, ATP and calmodulin.
- MISCELLANEOUS: The calcium release channel activity resides in the C-terminal region while the remaining part of the protein constitutes the 'foot' structure spanning the junctional gap between the SR and the T-tubule. It is possible that the foot structure interacts with the cytoplasmic region of the dihydropyridine receptor.
- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-release channel in junctional SR and modulates its activity.
- SIMILARITY: Belongs to the ryanodine receptor family.
- SIMILARITY: Contains 5 MIR domains.
- SIMILARITY: Contains 3 SPRY domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X62880; CAA44674.1; ALT\_SEQ.

EMBL; X68247; CAA48318.1; -.

EMBL; X69465; CAA49225.1; -.

EMBL; M32501; AAA31022.1; -.

InterPro: IPR000699; Ca-rel\_channel.

InterPro: IPR001682; Ca/Na\_pore.

InterPro: IPR002048; EF-hand.

InterPro: IPR005821; Ion\_trans.

InterPro: IPR003608; MIR.

InterPro: IPR001215; Ryanodn\_receptor.

InterPro: IPR003032; RYR.

InterPro: IPR003877; SERP\_receptor.

Pfam: PF00520; Ion\_trans; 1.

Pfam: PF02815; MIR; 4.

Pfam: PF01365; RYDR\_ITPR; 2.

Pfam: PF02026; RYR; 4.

Pfam: PF00622; SPRY; 3.

PRINTS: PR00795; RYANODINER.

SMART: SMO0472; MIR; 4.

SMART: SMO0449; SPRY; 3.

PROSITE: PS50919; MIR; 5.

Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat; Phosphorylation; Glycoprotein.

DOMAIN 1 3124

FT DOMAIN 1085 1208 SPRY 2.

FT DOMAIN 1430 1570 SPRY 3.

FT DOMAIN 1874 1925 GLU-RICH (ACIDIC).

FT DOMAIN 4458 4526 PRO-RICH.

FT DOMAIN 842 842 2960 6 X APPROXIMATE REPEATS.

FT REPEAT 842 955 1.

FT REPEAT 956 1069 2.

FT REPEAT 1345 1360 3 (INCOMPLETE).

FT REPEAT 1373 1388 4 (INCOMPLETE).

FT REPEAT 2227 2846 5.

FT REPEAT 2847 2960 6.

FT MOD\_RES 2844 2844 PHOSPHORYLATION (BY PKA AND PKG) (BY SIMILARITY).

FT MOD\_RES 3947 3947 PHOSPHORYLATION (POTENTIAL).

FT MOD\_RES 4320 4320 PHOSPHORYLATION (POTENTIAL).

FT CARBOHYD 3467 3467 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3475 3475 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3945 3945 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 4144 4144 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 4862 4862 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 2892 2092 A -> P (IN REF. 2 AND 3).

FT SEQUENCE 5035 AA; 565317 MW; E00613F2027B94A4 CRC64;

Query Match 4.4%; Score 85.5; DB 1; Length 5035;

Best Local Similarity 22.6%; Pred. No. 1.7e+02;

Matches 50; Conservative 24; Mismatches 74; Indels 73; Gaps 12;

Db 25 VPCGW-SRVSGIPSRHWP-VYKRFD--PRKPDYCOAYTCPTG----- 68

Db 1001 VAQGWSTYSAVODIPARRN-PLVPEYRLIDPATKRSNDSLCQAVRTLLGYNIPEPDQ 1059

Qy 69 -SPIPVNEGDIDIEVER-----LQAPVNEFKYDGLGLKIMDAIGFSTLTGKNTM 121

Db 1060 PSQVESQSRWDRKIFFAKXSVAVSGRWYFE-----FEAVTGE-MRV 1102

Qy 122 EWELEFQIGNCTPEHRLRPENDAPFWCQGAACPFEGIDVHWKENGITLVQATISGMFN 181

Db 1103 GWR-----PELRPDEL-----GDELALVYNGRGQRHLSGLF 1140

Qy 182 QMAKWKQD-----NETGIYETWNVKASPEKAEIWF 214

Db 1141 R--FWQSGDVGMIDITENTIIIFTLNGEVLMSDGSETA 1179

RESULT 12

SSA2 MOUSE STANDARD; PRT; 254 AA.

ID SSA2 MOUSE

AC 099N99;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE 3-oxo-5-alpha-steroid 4-dehydrogenase 2 (EC 1.3.99.5) (Steroid

DE 5-alpha-reductase 2) (SR type 2) (5 alpha-SR2).

GN SRD5A2 OR 5ART2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=21882004; PubMed=11884637;

RA Takeyama K., Kato S.;

RT "Transcriptional regulation of the mouse steroid 5alpha-reductase

RT type II gene by progesterone in brain."

RL Nucleic Acids Res. 30:1387-1393 (2002).

CC -1- FUNCTION: Converts testosterone into 5-alpha-dihydrotestosterone

CC and progesterone or corticosterone into their corresponding 5-

CC alpha-3-oxosteroids. It plays a central role in sexual

CC differentiation and androgen physiology (By similarity).

CC -1- CATALYTIC ACTIVITY: A 3-oxo-5-alpha-steroid + acceptor = a 3-oxo-

CC delta(4)-steroid + reduced acceptor.



```

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: TO OTHER MYCOBACTERIUM A85 ANTIGENS AND TO THE
CC N-TERMINAL OF C. GLUTAMICUM P81 PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D87323; BAA24161.1; -.
DR HSSP; P31953; IDQY.
DR InterPro; IPR000801; Esterase_pur.
DR Pfam; PF00756; Esterase; 1.
KM Transferase; Acyltransferase, Antigen, Signal.
FT SIGNAL 1 46
FT CHAIN 47 352
FT ACT SITE 170 170
FT ACT SITE 274 274
FT ACT_SITE 306 306
FT ACT_SITE 306 306
SQ SEQUENCE 352 AA; 37756 MW; A29277CB650D60A4 CRC64;

Query Match 4.3%; Score 83.5; DB 1; Length 352;
Best Local Similarity 23.4%; Pred. No. 8.8;
Matches 68; Conservative 26; Mismatches 123; Indels 73; Gaps 14;

QY 2 PRGGAGRGARSGWALALIMLAVPWSRVSGTFSRRHVPYRKRFDPKDPYCOAQ 61
DB 19 RLRLIAAVG-----ASLTSGVAVAAGGSPVAGAFSPGLVPEYLEY---PSSMGRIK 69
QY 62 YTFPPTSPSPY-----MEGDDIEVRLQAPWPEFYGGDLGKIMHAIFRSTL-- 114
DB 70 VQF-QGGGPHAVYLLDGLRAQDDINGNDINTPAREEPTSGSLVIMPVGGQSSFYSNWYQ 128
QY 115 ----TGKNTWMEWYELFQLGNTCFPHLRPEMDAPFWCNQGAACEFGIDVHMKENGTVL 170
DB 129 PSSGNGQNTYYTKM-ETP-----LTQEM-PLWMSKXQYSPAGNNAVGLSMGSGA 176
QY 171 QV-----ATISGMWENQMAKWKV---QDNETGTY--YETWNVKASPEKGAET 212
DB 177 LILAAVYPQGFPAASISGFLNPSEGMPPTLIGLANNDSGYNANSMGSPSTP-----A 231
QY 213 WEFSYDSCKFVLRTPENKLAEPGAEFPIKNIENYTRIFLY--SGEPTYLGN 260
DB 232 W-----KRNDPVQIPLVANNTRIWVCGNGTSPDLGD 266

RESULT 14
TM16 HUMAN
ID TM16 HUMAN STANDARD; PRT; 564 AA.
AC 095361; Q72612; Q96BB8; Q96U43;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tripartite motif protein 16 (Estrogen-responsive B box protein).
DE TRIM16 OR EBBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Vammary gland;
RX MEDLINE=99032537; PubMed=9817599;
RA Liu H.-L.C., Golder-Novoselsky E., Seto M.H., Webster L., McClary J.,
RA Zajchowski D.A.;
RT "The novel estrogen-responsive B box protein (EBBP) gene is tamoxifen
RT regulated in cells expressing an estrogen receptor DNA-binding domain
RT mutant.";
RT Mol. Endocrinol. 12:1733-1748(1998).
[2]

```

SEQUENCE FROM N.A. (ISOFORM 1).  
 MEDLINE=2037893; PubMed=11919186;  
 Beer H.-D., Munding C., Dubois N., Mamie C., Hohl D., Werner S.;  
 "The estrogen-responsive B box protein: a novel regulator of  
 keratinocyte differentiation.";  
 J. Biol. Chem. 277:20740-20749(2002).  
 [3]  
 SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 TISSUE=Lung, Placenta, and Skin;  
 MEDLINE=22388257; PubMed=12477932;  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner F.S., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Helel F.,  
 Diatchenko L., Maurina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.J.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,  
 Bosak S.A., McMan P.T., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalon D.K., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Butcherfield X.S., Krzyzinski M.I., Skalska U., Smalins D.E.,  
 Scherch A., Shein J.E., Jones S.J.M., Marra M.A.;  
 "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 -!- FUNCTION: May play a role in the regulation of keratinocyte  
 differentiation.  
 -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 -!- ALTERNATIVE PRODUCTS:  
 Event=Alternative splicing; Named isoforms=2;  
 Name=1; Synonyms=Alpha;  
 IsoId=O095361-1; Sequence=Displayed;  
 Name=2; Synonyms=Beta;  
 IsoId=O095361-2; Sequence=VSP\_009098;  
 -!- TISSUE SPECIFICITY: Highest levels found in testis, ovary, small  
 intestine, colon, placenta, heart, skeletal muscle and mammary  
 gland. More highly expressed in the fetus than in the  
 corresponding adult tissues. Expressed in basal keratinocytes.  
 -!- SIMILARITY: Belongs to the TRIM/RBCC family.  
 -!- SIMILARITY: Contains 1 B30.2-like domain.  
 -!- SIMILARITY: Contains 2 B box-type zinc fingers.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 EMBL; AF096870; AAC79080.1; -  
 EMBL; BC001564; AAH01564.1; -  
 EMBL; BC015674; AAH15674.1; -  
 EMBL; BC053514; AAH53514.1; -  
 Genew; HGNC:17241; TRIM6.  
 GO; GO:0005737; Cytoplasm; TAS.  
 InterPro; IPR001870; B302.  
 InterPro; IPR006574; PRY.  
 InterPro; IPR003877; SPRY receptor.  
 InterPro; IPR00315; Znf\_Box.  
 Pfam; PF006623; SPRY; 1.  
 Pfam; PF00643; Zf-B\_Box; 1.  
 PRINTS; PR01406; BBOXZNFINGER.  
 SMART; SM00336; BBOX; 2.  
 SMART; SM00589; PRY; 1.  
 SMART; SM00449; SPRY; 1.  
 PROSITE; PS00119; ZF\_BOX; 1.  
 Zinc-finger, zinc; Repeat; Coiled coil; Alternative splicing;

KW Polymorphism.  
 FT ZA FING 72 122 B BOX-TYPE 1.  
 FT ZA FING 126 165 B BOX-TYPE 2.  
 FT DOMAIN 165 203 COILED COIL (POTENTIAL).  
 FT DOMAIN 243 274 COILED COIL (POTENTIAL).  
 FT DOMAIN 320 340 COILED COIL (POTENTIAL).  
 FT DOMAIN 375 534 B30.2-LIKE.  
 FT VARSPPLIC 1 216 Missing (in isoform 2).  
 FT VARIANT 121 121 E -> D (in dbSNP:2074890).  
 FT CONFLICT 235 235 L -> P (IN REF. 1 AND 2).  
 FT CONFLICT 246 246 S -> N (IN REF. 1 AND 2).  
 FT CONFLICT 465 465 F -> L (IN REF. 3; AAH01564).  
 FT CONFLICT 521 521 K -> N (IN REF. 1 AND 2).  
 FT CONFLICT 541 541 N -> S (IN REF. 1 AND 2).  
 FT CONFLICT 561 561 V -> G (IN REF. 3; AAH01564/AAH15674).  
 SQ SEQUENCE 564 AA; 63997 MW; A676070D37D82F8 CRC64;  
 Query Match 4.3%; Score 83.5; DB 1; Length 564;  
 Best local similarity 22.1%; Pred. No. 16;  
 Matches 58; Conservative 31; Mismatches 81; Indels 93; Gaps 16;  
 QY 30 SRVSGIPSRHMPVVKRDPKPPD-----YCOAKY--TFCPGPSPIPVMEGDDIEV 82  
 DB 345 TVSAVAVQKRY-----TSKPEPSYREQFLQAYAYITDP-----DTAHKY 385  
 QY 83 FLQ-----APVPEKYGGLLHKMEALIGRSTLTGKN-YTMEWY--ELFQIG 130  
 DB 386 LRLQENRKYNTTWEHEHPDLPF--RPLH---WROVLQSOOSLYLHRYEVELEIFGAG 439  
 QY 131 NCTPHLHPENDAPFMCNCGACFEFGIDVHMKENGTLVQVATISGNMENGMAKVVXOD 190  
 DB 440 TVY-----GLTC--KGIDRKEERN-----SCISGNPFWSLQW---- 471  
 QY 191 NETGIYETW-----NYKASPEKGAETWFD-----SYDCKSEVLETPFKLAE- 232  
 DB 472 --NGKEFTAWYSDMETPLKAGFRLGVYIDPPGGLSPYGVETDMLVHFAKCFSESP 529  
 QY 233 -FGAFKRIETNYTPIFYSGSP 254  
 DB 530 VYAAFWLSKKEAIRIVDIGESP 552  
 RESULT 15  
 ID UROM\_HUMAN STANDARD; PRT; 640 AA.  
 AC P07911;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Uromodulin precursor (Tamm-Horsfall urinary glycoprotein) (THP).  
 GN UMOD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87177970; PubMed=3453112;  
 RA Pennica D., Kohr W.J., Kuang W.-D., Glaister D., Aggarwal B.B.,  
 Chen B.Y., Goeddel D.V.;  
 RT "Identification of human uromodulin as the Tamm-Horsfall urinary  
 glycoprotein";  
 RL Science 236:83-88(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=87319675; PubMed=3498215;  
 RA Hession C., Decker J.M., Sherblom A.P., Kumar S., Yue C.C.,  
 Mattaliano R.J., Tizard R., Kawashima E., Schneisener U.,  
 Heletsky S., Chow E.P., Burne C.A., Shaw A., Muchmore A.V.;  
 RT "Uromodulin (Tamm-Horsfall glycoprotein): a renal ligand for  
 lymphokines";



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on June 8, 2004, 13:59:08 ; Search time 63.9139 seconds

(without alignments)  
1708.068 Million cell updates/sec

File: US-10-010-050a-2

Sequence: 1927  
1 MRRGAGARGRASMCWALAL.....IKITYEIPILPIRKITLSGL 346

oring table: BL05UN62  
Gapop 10.0 , Gapept 0.5

arched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0  
imum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP plant:\*  
10: SP rodent:\*  
11: SP virus:\*  
12: SP vertebrate:\*  
13: SP unclassified:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacterioph:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1451	75.3	336	11 Q8R152	Q8R152 mus muscula
2	1429	74.2	322	11 Q8C054	Q8C054 mus muscula
3	1235	72.2	180	5 Q86157	Q86157 dictyosteli
4	202	10.5	378	5 Q86JG6	Q86JG6 dictyosteli
5	101.5	5.3	3118	5 Q17575	Q17575 caenorhabdi
6	101	5.2	633	3 Q8N1H1	Q8N1H1 trichophyto
7	99.5	5.2	514	16 Q8A3V2	Q8A3V2 bacteroides
8	98	5.1	682	4 Q8NB06	Q8NB06 homo sapien
9	97.5	5.1	804	5 Q8IAU5	Q8IAU5 plasmodium
10	96.5	5.0	1705	11 Q9ERK5	Q9ERK5 mus musculu
11	96	5.0	508	5 Q8IDS4	Q8IDS4 plasmodium
12	95.5	5.0	359	10 Q9ZV89	Q9ZV89 arabidopsis
13	95.5	5.0	836	4 Q13785	Q13785 homo sapien
14	95.5	5.0	5322	5 Q9VPL9	Q9VPL9 drosophila
15	94	4.9	362	5 Q9TYM2	Q9TYM2 caenorhabdi
16	94	4.9	1238	16 Q25330	Q25330 helicobacte

17	94	4.9	2313	8 Q8MHW9	Q8MHW9 psilocum nu
18	93	4.8	1349	16 Q8X6C1	Q8X6C1 escherichia
19	92.5	4.8	564	5 Q8M12	Q8M12 drosophila
20	92.5	4.8	569	5 Q9VEY2	Q9VEY2 drosophila
21	92	4.8	502	16 Q6561	Q6561 aquifex aeo
22	91	4.7	291	5 Q01702	Q01702 caenorhabdi
23	91	4.7	508	10 Q9XBC4	Q9XBC4 arabidopsis
24	91	4.7	607	4 Q9BVZ9	Q9BVZ9 homo sapien
25	91	4.7	742	4 Q94841	Q94841 homo sapien
26	91	4.7	742	4 Q723L6	Q723L6 homo sapien
27	90.5	4.7	850	2 Q8VUF6	Q8VUF6 azarcus ev
28	90.5	4.7	1846	5 Q814Z1	Q814Z1 plasmodium
29	90	4.7	1409	3 Q7Z918	Q7Z918 coprinus ci
30	89.5	4.6	444	10 Q93ZB3	Q93ZB3 arabidopsis
31	89.5	4.6	1512	2 Q9W4X5	Q9W4X5 streptococc
32	89	4.6	567	5 Q9TKX3	Q9TKX3 drosophila
33	89	4.6	570	5 Q8L1N6	Q8L1N6 drosophila
34	89	4.6	584	5 Q8MT35	Q8MT35 drosophila
35	89	4.6	613	5 Q9M122	Q9M122 drosophila
36	89	4.6	1353	5 Q9XK65	Q9XK65 caenorhabdi
37	88.5	4.6	465	10 Q9ZS50	Q9ZS50 ipomoea bat
38	88.5	4.6	540	2 Q9A1P9	Q9A1P9 methylococc
39	88.5	4.6	752	17 Q9YC61	Q9YC61 aeropyrum p
40	88.5	4.6	1462	4 Q86U07	Q86U07 homo sapien
41	88	4.6	438	3 Q7Z9M6	Q7Z9M6 trichoderma
42	88	4.6	661	12 Q9JG50	Q9JG50 ttv-like mi
43	88	4.6	742	6 Q29397	Q29397 bos taurus
44	88	4.6	2107	12 Q91NQ2	Q91NQ2 tauro syndr
45	88	4.6	2713	5 Q8MHW0	Q8MHW0 dictyosteli

#### ALIGNMENTS

RESULT 1  
ID Q8R152 PRELIMINARY: PRT: 336 AA.  
AC Q8R152.  
DT 01-JUN-2002 (TRENBLER, 21, Created)  
DT 01-JUN-2002 (TRENBLER, 21, Last sequence update)  
DT 01-JUN-2003 (TRENBLER, 24, Last annotation update)  
DE Similar to ceroid-lipofuscinosis, neuronal 5 (Fragment).  
GN C1N5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; RC025487; AAH25487.1; -;  
DR MGD; MGI:2442253; C1N5.  
FT NON TER  
SQ SEQUENCE 336 AA; 38911 MW; D50D74C45483BE1B CRC64;

Query Match 75.3%; Score 1451; DB 11; Length 336;  
Best Local Similarity 76.6%; Pred. No. 1.6e-118;  
Matches 258; Conservative 31; Mismatches 42; Indels 6; Gaps 2;

QY	12	ASW--CVALALMLAVVGVGSRVSGIPSRHHMPVYKRPFRPPDPYCOAKYFCEPTGS	69
QY	4	AHMRPALALMLGLATLGLASPTSG---GRWPVYKRSFRPKTDYCOAKYFCEPTGS	59
QY	70	PIPYMGDDIEVFRLLQAPVWEFRYGDLLGLTKIMDAIGFRSTLTGKNYTMWYELFOL	129
QY	60	PIPYMKNDVIEVLRIGAPVWEFRYGDLLGLTKIMDAVGRSTLTGKNYTMWYELFOL	119
QY	130	GNCTFPIRPEMDAPFVWNGGAACFFEGIDVAKKNGTIVQVATIGKKNFNQAKAVKQ	189
QY	120	GNCTFPIRPEMDAPFVWNGGAACFFEGIDVAKKNGTIVQVATIGKKNFNQAKAVKQ	179
QY	190	DNETGIYETWNVKASPEKGAETWFDSDSKFVLRTFNKLAEGAFKNIETVYTIPL	249





```

Y 228 NKLAEFGAEFFKQLEITNYTRIFLYSGEPTLYGNETSVFQGTGKLTGLAIKRFYEPKPHL 287
b 293 NMLAQNN-----FDGGEEDLYNN-----YRP-NSKNL-----KREYFAPRM 327
Y 288 -PKREFL-LSLQIF-DAVIYAKQFL 311
b 328 SPEKSYIDASVTLQLFYSANIVHDLHYM 354

RESULT 7
Q8A3V2 PRELIMINARY; PRT; 514 AA.
C Q8A3V2;
T 01-JUN-2003 (TrEMBLrel. 24, Created)
T 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Putative beta-xylosidase.
N BT2852.
S Bacteroides thetaiotaomicron.
K Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
K Bacteroidaceae; Bacteroides.
X NCBI_TaxID=818;

[1]
SEQUENCE FROM N.A.
C STRAIN=VPI-5482 / ATCC 29148;
C MEDLINE=22550858; PubMed=12663928;
A Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
A Chiang H.C., Hooper L.V., Gordon J.I.,
A "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
L Science 299:2074-2076(2003).
R EMBL; AB016937; AA077958.1;
R GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
R GO; GO:0005975; P:carbohydrate metabolism; IEA.
R InterPro; IPR006710; Glyco_hydro_43.
R Pfam; PF04616; Glyco_hydro_43; 1.
M Complete proteome.
X SEQUENCE 514 AA; 58155 MW; 6F6E8653CC9488 CRC64;

Query Match 5.2%; Score 99.5; DB 16; Length 514;
Best Local Similarity 21.8%; Pred. No. 3.7;
Matches 52; Conservative 30; Mismatches 92; Indels 65; Gaps 9;

Y 41 WPPVYKRFDPFRPPDPYCAKYTCFPTGSPDPVM--EGDDIEVFRLQAVWEFKYGD-- 96
b 294 WPIYGVIDIMNGGEPY--KVTIKPTNGKKVPSFGGSDSPSPSLNIQ--WGFNNPSD 350
Y 97 -----LIGHIKTME-----DAIGFRSTLTGKNYTWMEY-----LPOL 129
b 351 ADWNLTERKGMILLKALKADHLRASHNMLTQKCIYEYGVITTEMDSWTEGQAGIFCI 410
Y 130 GNCCTPHLRPMADPFWNGGAACFPGI-----DVHMKENGTLVQVATISGMFN 181
b 411 GN-----LENGIGILKENGKNYLYLNNNSVEKVKSGKTY 448
Y 182 QMAKWKQDNETGIYETWNVKASP-EKGAETWFDSDYSCFKYLRTFNKLAEFGAEFKN 239
b 449 FRATMNAKRTQHLYYSTDKNTTPCEAASLRFGMKGRVGLYSNTLRDGNAPFN 507

RESULT 8
Q8A3V2 PRELIMINARY; PRT; 682 AA.
C Q8A3V2;
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E Hypothetical protein NT2RP2002760.
S Homo sapiens (human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
X NCBI_TaxID=9606;
M [1]
SEQUENCE FROM N.A.

```

```

RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA Nagahari K., Sugano S., Isogai T.;
RT "HRT human cDNA sequencing project.";
CC Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
C -i- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
R EMBL; AK075480; BAC11645.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub-transporter.
DR InterPro; IPR005829; Sug transporter.
DR InterPro; IPR005988; SV2.
DR Pfam; PF00083; sugar tr; 1.
DR TIGRFAMs; TIGR01259; synapt_SV2; 1.
DR PROSITE; PS00850; MFS; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transmembrane.
X SEQUENCE 682 AA; 76660 MW; DIA76C11542ACC7F CRC64;

Query Match 5.1%; Score 98; DB 4; Length 682;
Best Local Similarity 22.1%; Pred. No. 7.3;
Matches 62; Conservative 40; Mismatches 104; Indels 74; Gaps 16;

```

```

Y 116 GKNYTWMEYELFQNGCTEPH-----LRPMDAP-FMCGQ-----AACFEGIDDVHMK 164
b 325 GSAYGFHSRVRVFL-VCAEPVSFALGALTQEPSEPRFLFNKHDEAMVVLQVHDTNMR 383
Y 165 ENGLTVQATISG-NMFMQMAKWKQDNETGIYETWNVKASPEKGAETWFDSDYSCKFV 223
b 384 AGKHGERVPSVNIHKTIHQEGELIEQSDPTGYWRMGVRAL-SLGGQVWGNFLSC----- 438
Y 224 LATFNKLAERGAEPFNIEFN-----YTRIFLYSGERTYL-----GNISYVF-- 264
b 439 -----FGEYRRTILMMGWFTWSFSYGGTLTWPPDMIRHLQAVDVASRTVFPFG 489
Y 265 -----GPTGKTKTGLAIKRFYEPK-----HLPTEFLSLIQ--IFDAVIYAKQFY-- 310
b 490 EKEVHTFETFLNQHNGQYFNDFKGLRKSASFBSLSEBCEFEVLTSSNAFRRNC 549
Y 311 -----LEFN--FEYWFELPKRPFKITY-----EELPDI 338
b 550 TEINTEFYNTDLFEYKFNVS--LINSITLHNKEDCELDV 587

RESULT 9
Q8IAU5 PRELIMINARY; PRT; 804 AA.
C Q8IAU5;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E Cullin-1-like protein, putative.
N PF08_0094.
S Plasmodium falciparum (isolate 3D7).
C Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
X NCBI_TaxID=36329;
M [1]
SEQUENCE FROM N.A.
R Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
R Quail M., Barrrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
R EMBL; AL844507; CAD51265.1; -.
DR GO; GO:0007049; P:cell cycle; IEA.
DR InterPro; IPR001373; Cullin.
DR Pfam; PF00888; Cullin; 1.
DR SMART; SM00182; CULLIN; 1.
DR PROSITE; PS50069; CULLIN_2; 1.
X SEQUENCE 804 AA; 97162 MW; 6B6E916BA99EAD2C CRC64;

Query Match 5.1%; Score 97.5; DB 5; Length 804;
Best Local Similarity 19.8%; Pred. No. 10;

```



	Matches	57,	Conservative	51,	Mismatches	115,	Indels	65,	Gaps	14,
y	53	KPDYQCAKTKFECTSGSPI-----	PMEGDDDIEVFLQAPWMEKRY-----	GD	96					
b	57	KKNPFCSYKSEYRRKYGSLSMYTTIDIKIKPLKNSDELNTKTLIDAM-KFYSFYTKMANK		115						
y	97	LLGHILKMHDAIGRSTLTGKNYTMMEVELFQLGNCFTPHLIPBMDAPWCNQAACFE		156						
b	116	FLRIADRY--VEVNSSLCISAVTKNIFKI-----	ILTFBELR-----	15..						
y	157	GIDVDEHK-ENGTLVQVATISGMNEMAKWVQ-DNEG-IYYE-----	TNRYKASP	206						
b	152	DIKAIYEIYNNLFLQELIDQKELFCNCIVAEIYELNELNESNEKYEHDIKKIVENVNNFY		21..						
y	207	EKGAEIWFSDYSCSKFLRTFNKLAEFGAEFFNIETN-----	YTRIFLYSGEPTY	256						
b	212	KKKEEWINDYPPDDYITISENAIKEKEYEKNKSLINDNPTCEKVTNIIVKILITRYELMTL		27..						
y	257	LGNTSVGPTGNKTLGLAIKRFPYPPKHLPTKEELLSILQIFDAVI		304						
y	272	IDNKNNIFHLIKNNNIS-SLRRTYILPS-YEP--EKLGTGKIKIGEYI		315						

ESULT 10  
9ERKS  
D Q9ERKS PRELIMINARY; PRT; 1705 AA.

T	01-MAR-2011 (TREMblrel. 16, Created)
T	01-MAR-2001 (TREMblrel. 15, Last sequence update)
T	01-OCT-2003 (TREMblrel. 26, Last annotation update)
E	Osteoesticular protein tyrosine phosphatase (EC 3.1.3.403).
N	PTPR OR ESP.
S	Mus musculus (Mouse).
C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
X	Mammalia; Eubacteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X	NCBI_TaxID=10090;
N	[1]
P	SEQUENCE FROM N.A.
C	TISSUE=Bone;
C	MEDLINE=20534790; PubMed=11060566;
A	Morrison D.F., Mauro L.J.;
T	"Structural characterization and chromosomal localization of the mouse
T	cDNA and gene encoding the bone tyrosine phosphatase, MOST-PTP.";
L	Gene 257:195-208(2000).
R	EMB1; AF300701; AAC82768.1; -.
R	HSSP; P18052; 1YFO.
R	MGD; MGI:108027; Ptdrv.
R	GO; GO:0016787; F:hydrolyase activity; IEA.
R	GO; GO:0004727; F:phosphatase activity; IEA.
R	GO; GO:0006470; F:protein amino acid dephosphorylation; IEA.
R	InterPro; IPR003961; FN.II1.
R	InterPro; IPR008957; FN.II1-like.
R	InterPro; IPR00387; Tyr Phosphatase.
R	InterPro; IPR00242; Tyr_PP.
R	Pfam; PF00041; fn3; 7.
R	Pfam; PF00102; Y_phosphatase; 1.
R	PRINTS; PR00700; PRTYPHPTASE.
R	SMART; SM00060; FN3; 7.
R	SMART; SM00194; PTPc; 1.
R	PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
R	PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
R	PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
R	Hydrolase.
Q	SEQUENCE 1705 AA; 186810 MW; 84A8EFA812EB53C1 CRC64;
Q	SEQUENCE
Q	Best Local Similarity 5.0%; Score 96.5; DB 11; Length 1705;
Q	Matches 78; Conservative 40; Mismatches 134; Indels 113; Gaps 17
Q	17
Q	39
Q	88

```

QY 40 -----HMPYKRFD-----FRPKRPYQAKYFC-----PGSGT 71
Db 89 NESSFEFFHGLPGSGRYQJELTVLP--CWQNVITLTAAFTAVYVGLQJLHSGSPA 143
QY 72 FV-----MEGDDIEVERLQAP-----VWEFKYGBLLGH.KJTMHDAIGR 111
Db 144 SLEMSWDSASDDQDSYQJLLYHPESHTLACVSVSPDLSYVFGJLLPQSQYVLEVITTA 203
QY 112 STLKKNYTWMEWYELFQJGNCTFPHLRBEMDAPEW-CNQACAFPEGJDDVHWKENGTLV 170
Db 204 GSLHAKNSILQMTPEVPDPDHLTLRATGTSJQAFNNSBGATWFIHLLDLDLGNLTJVK 263
QY 171 QVATISGNMENQMAKWKVQKQNETGIY-----ETANVKASPEKGAETWFDSYDC 219
Db 264 VRQGISHTHTFRLS-----PGRPYQJ.KICAAAPHQIWMPNAT-----EMTYSYPS 310
QY 220 SKFYLRFTNKLAEGBAEFKNIETNYTRLFYSG--EPTYLGNF--TSYFG--PTGNMT 271
Db 311 DLVYTPPLMNEELMASWKAGQARGDY-VLKLSGPVENTTTLGPEBCNAVFPGBLPPGHYT 368
QY 272 LGLAI 216
Db 369 LGLRV 373

```

RESULT 11	
Q8IDS4	
ID Q8IDS4	PRELIMINARY; PRT; 508 AA.

DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF13\_0225.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
 RA Bertram M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Omond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL044509; CAD52545.1; -.  
 DE Hypothetical protein.  
 QX SEQUENCE 508 AA; 58842 MW; F563E34C1A9F25D2 CRC64;

Query Match	5.0%;	Score 96;	DB 5;	Length 508;
Best Local Similarity	20.0%;	Pred. No. 7.4;		
Matches	52;	Conservative	33;	Mismatches 95; Indels 80; Gaps 11.

```

QY      156  GAGIDDVHKEKGTLYQVAVTISGNNEQKAKWKOENEGIYYE-----TAVKASPEKG  209
Db      121  DNINNNMHHNNNNFNVTTLDNRKFMKSKSNMNDPSNITTYMTLNNDGINMKKPYPSKI  180
QY      210  AETWFDSDYCSKFEVLRTFN-----KLAEGFAEFKNIETN-YTRIPLYS  251
Db      181  EN--YDYVDIENFPRQTNNOQLYSANNIVPLGYRNYKIQENGDSINGYATNTYKIGHGHTV  238
QY      252  GE-----PNTL-----GNETSVAGPGTKTL  272
Db      239  GRIKPFNVLGACQLESLCTPSYTTISNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNILGFPENKTL  298
QY      273  GLAI-----RRFYFPRKPHLPYKEFELLSTLQIPDAVIVHKQPYLDFYNEFWYFLPMKPEF  326
Db      299  NMPYTLNDQNNKEYIIPY-----LNKGMEITNDYMMGMSYTNK--FSPNVEYELMTNDRSPY  351
QY      327  IKITY-BEIEPLPRNK-TLS  344
Db      352  YMLNNTYNEKLDIFPKITLS  371

```

## RESULT 12





de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,  
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
 Jalali M., Kainush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svitskas R., Tecor C., Turner R., Venter F., Wang A.H., Wang X.,  
 Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 "The genome sequence of *Drosophila melanogaster*."  
 Science 287:2185-2195 (2000).  
 [2]  
 SEQUENCE FROM N.A.  
 Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 Barzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 Dodson K., Dorsetti V., Doup L.E., Doyle C., Dresek D., Farfan D.,  
 Ferrieri S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 Ibeagwam C., Jalali M., Kruse D., Li P., Matel B., Moshrefi A.,  
 McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 Phouaneavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 Stapleton M., Strong R., Svitskas R., Tecor C., Tyler D.,  
 Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 "Sequencing of *Drosophila melanogaster* genome."  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 SEQUENCE FROM N.A.  
 Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celinker S.E.,  
 Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 "Annotation of *Drosophila melanogaster* genome."  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 SEQUENCE FROM N.A.  
 Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 SEQUENCE FROM N.A.  
 FlyBase;  
 Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 [6]  
 SEQUENCE FROM N.A.  
 Therrien M., Morrison D.K., Wong A.M., Rubin G.M.,  
 "A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype  
 in *Drosophila*."  
 Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 L. EMBL; AF005590; AA51527.3; -  
 R. EMBL; AF215703; AA543004.1; -  
 R. HSSP; P23197; IAP0.  
 R. FlyBase; FBgn0001309; kis.  
 R. GO; GO:000785; C:chromatin; IEA.

DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003682; F:chromatin binding; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.  
 DR InterPro; IPR000953; Chromo.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR Pfam; PF00385; chromo; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR PROSITE; PS50013; CHROMO\_2; 1.  
 DR ATP-binding; Helicase; Hydrolase.  
 SQ SEQUENCE 5322 AA; 573615 MW; C9608375FA71C211 CRC64;

Query Match 5.0%; Score 95.5; DB 5; Length 5322;

Best Local Similarity 20.1%; Pred. No. 1.8e+02; Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;

31 RVSGIPSRHMPVYKPRPKDPYCAKTCPTGSPFVMEGDDIEVFRLOAPVW 90  
 1993 RPNKIPQSEW-----KSKKPHPELMKKLEKTE-----PVYKGSNLRPYQLEGNW 2039

91 EFKYDDLGLKLMNDAGFSTLTGKNTYMEWYELFOLGNTFPHLPKMAPWNCOG 150  
 2040 LKFSWYNTNHCILADEMGLGTLTSLFVSVYEGIRG--PFLVIAFLSTIPWQRE- 2095

151 AACFREGIDVHWKNGTLVQATISGMFMNKAQWQODNETGYETPMWVKASPE--- 207  
 2096 ---PEG---WIDMNVVYHGSVT-----SKQMLQDY-YYNTESKVLKEPIK 2137

208 KGAETWDSYDCSKFVLRTEF-----KLAEGAEKNIE----- 241  
 2138 FNVLLTFEMVLYTMDKAFWRMLCVIDEAHRLLKRNCKLLE-GLRLNLEHRYLST 2196

242 ---TNYRIF-LYSGEPYVGNETSVPFGPTGNKTLGLAKFPYEPFHL----- 287  
 2197 PLQNNISLFLSNLFLFESQSFESGFLTEEBVNLQALLKPMRLRLKDVE 2256

288 ---PTKEFLSLDLPDAVYHKQFLFYNEVWFLPMKKEPIK--ITYEELP 335  
 2257 KSLAKETTLLEV---ELTNLQKRY-----RGLPEONFSLKKGTTSANIP 2300

RESULT 15  
 ID Q9TYM2 PRELIMINARY; PRT; 362 AA.

AC Q9TYM2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Y25C1A.7a protein.  
 GN Y25C1A.7 OR Y25C1A.7A.  
 OS Caenorhabditis elegans.

OC Caenorhabditis elegans.  
 OC Rhabditidae; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., McKurray A., Koopra A., Saunders D., Showkhen R.,

```

1A Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
2A Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
3A Watson A., Weinstock L., Wilkinson-Spratt J., Woldman P.;
4A "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
5A elegans.";
6A Nature 368:32-38(1994).
7A
8A [2]
9A SEQUENCE FROM N.A.
10A STRAIN=Bristol N2;
11A Kalicki J., Smith A., Gibson A.;
12A "The sequence of C. elegans cosmid Y25C1A.";
13A Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
14A [3]
15A SEQUENCE FROM N.A.
16A STRAIN=Bristol N2;
17A Waterston R.;
18A Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
19A EMBL; AF125459; AAD12838.1; -.
20A PIR; T33904; T33904.
21A WormPep; Y25C1A.7a; CE21474.
22A InterPro; IPR006977; DUF649.
23A Pfam; PF04893; Yip1; 1.
24A SEQUENCE 362 AA; 40613 MW; 42DB45A7F72F3B89 CRC64;
25A
26A
27A
28A
29A
30A
31A
32A
33A
34A
35A
36A
37A
38A
39A
40A
41A
42A
43A
44A
45A
46A
47A
48A
49A
50A
51A
52A
53A
54A
55A
56A
57A
58A
59A
60A
61A
62A
63A
64A
65A
66A
67A
68A
69A
70A
71A
72A
73A
74A
75A
76A
77A
78A
79A
80A
81A
82A
83A
84A
85A
86A
87A
88A
89A
90A
91A
92A
93A
94A
95A
96A
97A
98A
99A
100A
101A
102A
103A
104A
105A
106A
107A
108A
109A
110A
111A
112A
113A
114A
115A
116A
117A
118A
119A
120A
121A
122A
123A
124A
125A
126A
127A
128A
129A
130A
131A
132A
133A
134A
135A
136A
137A
138A
139A
140A
141A
142A
143A
144A
145A
146A
147A
148A
149A
150A
151A
152A
153A
154A
155A
156A
157A
158A
159A
160A
161A
162A
163A
164A
165A
166A
167A
168A
169A
170A
171A
172A
173A
174A
175A
176A
177A
178A
179A
180A
181A
182A
183A
184A
185A
186A
187A
188A
189A
190A
191A
192A
193A
194A
195A
196A
197A
198A
199A
200A
201A
202A
203A
204A
205A
206A
207A
208A
209A
210A
211A
212A
213A
214A
215A
216A
217A
218A
219A
220A
221A
222A
223A
224A
225A
226A
227A
228A
229A
230A
231A
232A
233A
234A
235A
236A
237A
238A
239A
240A
241A
242A
243A
244A
245A
246A
247A
248A
249A
250A
251A
252A
253A
254A
255A
256A
257A
258A
259A
260A
261A
262A
263A
264A
265A
266A
267A
268A
269A
270A
271A
272A
273A
274A
275A
276A
277A
278A
279A
280A
281A
282A
283A
284A
285A
286A
287A
288A
289A
290A
291A
292A
293A
294A
295A
296A
297A
298A
299A
300A
301A
302A
303A
304A
305A
306A
307A
308A
309A
310A
311A
312A
313A
314A
315A
316A
317A
318A
319A
320A
321A
322A
323A
324A
325A
326A
327A
328A
329A
330A
331A
332A
333A
334A
335A
336A
337A
338A
339A
340A
341A
342A
343A
344A
345A
346A
347A
348A
349A
350A
351A
352A
353A
354A
355A
356A
357A
358A
359A
360A
361A
362A
363A
364A
365A
366A
367A
368A
369A
370A
371A
372A
373A
374A
375A
376A
377A
378A
379A
380A
381A
382A
383A
384A
385A
386A
387A
388A
389A
390A
391A
392A
393A
394A
395A
396A
397A
398A
399A
400A
401A
402A
403A
404A
405A
406A
407A
408A
409A
410A
411A
412A
413A
414A
415A
416A
417A
418A
419A
420A
421A
422A
423A
424A
425A
426A
427A
428A
429A
430A
431A
432A
433A
434A
435A
436A
437A
438A
439A
440A
441A
442A
443A
444A
445A
446A
447A
448A
449A
450A
451A
452A
453A
454A
455A
456A
457A
458A
459A
460A
461A
462A
463A
464A
465A
466A
467A
468A
469A
470A
471A
472A
473A
474A
475A
476A
477A
478A
479A
480A
481A
482A
483A
484A
485A
486A
487A
488A
489A
490A
491A
492A
493A
494A
495A
496A
497A
498A
499A
500A
501A
502A
503A
504A
505A
506A
507A
508A
509A
510A
511A
512A
513A
514A
515A
516A
517A
518A
519A
520A
521A
522A
523A
524A
525A
526A
527A
528A
529A
530A
531A
532A
533A
534A
535A
536A
537A
538A
539A
540A
541A
542A
543A
544A
545A
546A
547A
548A
549A
550A
551A
552A
553A
554A
555A
556A
557A
558A
559A
560A
561A
562A
563A
564A
565A
566A
567A
568A
569A
570A
571A
572A
573A
574A
575A
576A
577A
578A
579A
580A
581A
582A
583A
584A
585A
586A
587A
588A
589A
590A
591A
592A
593A
594A
595A
596A
597A
598A
599A
600A
601A
602A
603A
604A
605A
606A
607A
608A
609A
610A
611A
612A
613A
614A
615A
616A
617A
618A
619A
620A
621A
622A
623A
624A
625A
626A
627A
628A
629A
630A
631A
632A
633A
634A
635A
636A
637A
638A
639A
640A
641A
642A
643A
644A
645A
646A
647A
648A
649A
650A
651A
652A
653A
654A
655A
656A
657A
658A
659A
660A
661A
662A
663A
664A
665A
666A
667A
668A
669A
670A
671A
672A
673A
674A
675A
676A
677A
678A
679A
680A
681A
682A
683A
684A
685A
686A
687A
688A
689A
690A
691A
692A
693A
694A
695A
696A
697A
698A
699A
700A
701A
702A
703A
704A
705A
706A
707A
708A
709A
710A
711A
712A
713A
714A
715A
716A
717A
718A
719A
720A
721A
722A
723A
724A
725A
726A
727A
728A
729A
730A
731A
732A
733A
734A
735A
736A
737A
738A
739A
740A
741A
742A
743A
744A
745A
746A
747A
748A
749A
750A
751A
752A
753A
754A
755A
756A
757A
758A
759A
760A
761A
762A
763A
764A
765A
766A
767A
768A
769A
770A
771A
772A
773A
774A
775A
776A
777A
778A
779A
780A
781A
782A
783A
784A
785A
786A
787A
788A
789A
790A
791A
792A
793A
794A
795A
796A
797A
798A
799A
800A
801A
802A
803A
804A
805A
806A
807A
808A
809A
810A
811A
812A
813A
814A
815A
816A
817A
818A
819A
820A
821A
822A
823A
824A
825A
826A
827A
828A
829A
830A
831A
832A
833A
834A
835A
836A
837A
838A
839A
840A
841A
842A
843A
844A
845A
846A
847A
848A
849A
850A
851A
852A
853A
854A
855A
856A
857A
858A
859A
860A
861A
862A
863A
864A
865A
866A
867A
868A
869A
870A
871A
872A
873A
874A
875A
876A
877A
878A
879A
880A
881A
882A
883A
884A
885A
886A
887A
888A
889A
890A
891A
892A
893A
894A
895A
896A
897A
898A
899A
900A
901A
902A
903A
904A
905A
906A
907A
908A
909A
910A
911A
912A
913A
914A
915A
916A
917A
918A
919A
920A
921A
922A
923A
924A
925A
926A
927A
928A
929A
930A
931A
932A
933A
934A
935A
936A
937A
938A
939A
940A
941A
942A
943A
944A
945A
946A
947A
948A
949A
950A
951A
952A
953A
954A
955A
956A
957A
958A
959A
960A
961A
962A
963A
964A
965A
966A
967A
968A
969A
970A
971A
972A
973A
974A
975A
976A
977A
978A
979A
980A
981A
982A
983A
984A
985A
986A
987A
988A
989A
990A
991A
992A
993A
994A
995A
996A
997A
998A
999A
1000A

```

search completed: June 8, 2004, 14:07:04  
Job time : 66.9139 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 13:55:43 ; Search time 7.42778 Seconds  
(without alignments)  
1065.101 Million cell updates/sec

File: US-10-010-050a-2\_copy\_1\_28  
Project score: 154  
Sequence: 1 MRRGAGARGRASMCWMLALIMLAVPG 28

oring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0  
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	347	2	AAV41323	AAV41323 Human sec
2	154	100.0	358	4	AAW93870	AAW93870 Human pol
3	147	95.5	346	2	AAW92967	AAW92967 Human zbl
4	60.5	39.3	192	4	AAU29418	AAU29418 Human g p
5	60.5	39.3	192	5	ABG60706	ABG60706 Novel g p
6	59.5	38.6	127	4	AAU14162	AAU14162 Human nov
7	58	37.7	188	5	ABU60973	ABU60973 Lung spec
8	57.5	37.3	1938	6	ABP76682	ABP76682 Streptomy
9	56	36.4	214	5	ABU05547	ABU05547 M. tuberc
10	55	35.7	403	5	ABU65190	ABU65190 Human NOV
11	54	35.1	78	4	ABU16658	ABU16658 Human ner
12	53	34.4	34	2	AAV43189	AAV43189 Thioeredox
13	53	34.4	211	4	ABG01709	ABG01709 Novel hum
14	53	34.4	346	5	ABG96249	ABG96249 Matize per
15	53	34.4	600	4	ABG07764	ABG07764 Novel hum
16	52.5	34.1	199	4	AAU28370	AAU28370 Novel hum
17	52.5	34.1	199	7	ADU09107	ADU09107 Novel pro
18	52.5	34.1	199	7	ADU09108	ADU09108 Novel pro
19	52.5	34.1	267	4	AAW79568	AAW79568 Corynebact
20	52.5	34.1	333	3	AAW91017	AAW91017 C. glutami
21	52.5	34.1	341	3	AAW69880	AAW69880 B. lactofi
22	52.5	34.1	360	6	ABU22472	ABU22472 Protein e
23	52.5	34.1	442	1	AAW90475	AAW90475 N-termina
24	52.5	34.1	551	2	AAW76503	AAW76503 Cattle MI
25	52.5	34.1	575	1	AAW70195	AAW70195 Sequence

## ALIGNMENTS

RESULT 1	AAV41323	standard; protein; 347 AA.
XX	AAV41323;	
AC	AAV41323;	
DT	02-DEC-1999 (first entry)	
XX	Human secreted protein encoded by gene 16 clone HMZAD77.	
XX	Human, secreted protein; fusion protein; gene therapy; protein therapy;	
XX	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;	
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;	
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KW	inflammation; ischaemic shock; Alzheimer's disease; osteoclast; thymus;	
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
XX	OS Homo sapiens.	
XX	WO9947540-A1.	
PN	23-SEP-1999.	
XX	18-MAR-1999;	99WO-US005804.
PF	19-MAR-1998;	98US-0078563P.
XX	19-MAR-1998;	98US-0078566P.
PR	19-MAR-1998;	98US-0078573P.
PR	19-MAR-1998;	98US-0078574P.
PR	19-MAR-1998;	98US-0078576P.
PR	19-MAR-1998;	98US-0078577P.
PR	19-MAR-1998;	98US-0078578P.
PR	19-MAR-1998;	98US-0078579P.
PR	19-MAR-1998;	98US-0078581P.
PR	01-APR-1998;	98US-0080312P.
PR	01-APR-1998;	98US-0080313P.
PR	01-APR-1998;	98US-0080314P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;	
PI	Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Latleur DW, Olsen HS;	
PI	Shi Y, Moore PA,	
XX	WPI, 1999-562050/47.	
DR	N-PSDB; AA24826.	

	TS	New isolated human genes, useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.
	TU	
	TV	
	TX	Claim 11; Page 367-368; 484pp; English.
	TY	This sequence represents a secreted human protein encoded by the gene clone detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AA224602) for increasing the stability of the fused protein as compared to the human protein only. The invention relates to 95 novel genes and their fragments (nucleic acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 95 polynucleotides, based on which tissues they are most highly expressed in (see AA224811 for described uses)
	XZ	Sequence 347 AA;
	YJ	Query Match Best Local Similarity 100.0%; Score 154; DB 2; Length 347; Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0
	ZB	1 MRGGAAGRGRASWCMALATLMLAVPG 28     1 MRGGAGARGRASWCMALATLMLAVVG 28
	AA	RESULT 2 LAM93870 ID AAM93870 strand; protein; 358 AA.
	AC	AAM93870;
	AD	06-NOV-2001 (first entry)
	AE	Human polypeptide, SEQ ID NO: 3978.
	AF	Human; full length cDNA; cDNA synthesis; oligo-capping.
	AG	Homo sapiens.
	AH	EP1130094-A2.
	AI	05-SEP-2001.
	AJ	07-JUL-2000; 2000EP-00114089.
	AK	08-JUL-1999; 99JP-00194486.
	AL	11-JAN-2000; 2000JP-00118774.
	AM	02-MAY-2000; 2000JP-00183765.
	AN	(HELI-) HELIX RES INST.
	AO	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y, Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H; WPI; 2001-524255/58. N-FSDB; AAK94829.
	AP	830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation.
	AQ	Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.
	AR	The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful.

	CC	for clarifying the function of the protein encoded by the cDNA. The full
	CC	length clones were obtained by construction of full length enriched cDNA
	CC	libraries that were synthesized by the oligo-capping method. The primers
	CC	enable the production of the full length cDNA easily without any special
	CC	methods. The present sequence is a polypeptide encoded by a full length
	CC	human cDNA of the invention. Note: The sequence data for this patent did
	CC	not form part of the printed specification, but was obtained in CD-ROM
	CC	format directly from EPO
XX	SQ	Sequence 358 AA;
XY	Query Match	100.0%; Score 154; DB 4; Length 358;
Ddb	Best Local Similarity	100.0%; Pred. No. 3 2e-13;
	Matches	28; Conservative 0; Mismatches 0; Indels 0; Gaps 0
		1 MRRGAGARGRASWCWALATLMTLAIVVPG 28
		13 MRGAGARGRASWCWALATLMTLAIVVPG 40
RESULT 3		
ID AAM92967	standard; protein; 346 AA.	
AC AAM92967;		
DT 14-MAY-1999	(first entry)	
DE Human zsig46 protein.		
KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;		
KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;		
KW Hirschsprung's disease; neuronal ceroid-lipofunctosis; Wilson disease;		
KW Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;		
KW therapy; diagnostic.		
OS Homo sapiens.		
PN MO9905275-AI.		
PD 04-FEB-1999.		
Pf 24-JUL-1998;	98WO-US015431.	
PR 24-JUL-1997;	97US-0053613P.	
PA (ZYMO ) ZYMOGENETICS INC.		
PI Sheppard PO, Gilbertson DG;		
XX WPI; 1999-142930/12.		
DR N-PSDB; AAX02855.		
FT New secreted polypeptide, zsig46, and its fragments, related fusion		
PT proteins - used for diagnosis and treatment of thyroid disorders or		
XX diseases involving genes on chromosome 13.		
PS Claim 3; Page 90-91; 101pp; English.		
CC This invention describes the isolation of a novel human secreted protein,		
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in from		
CC the thyroid. This product can be used to study secretion of zsig46, which		
CC cells and also to treat or prevent deficient expression of zsig46, which		
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'		
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that		
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's		
CC disease, neuronal ceroid-lipofunctosis, Wilson disease and Reiger		
CC syndrome). Antibodies and other binding proteins, are used as immunoassay		
CC reagents to detect zsig46 or cells expressing it, e.g. for assessing		
CC thyroid function to produce anti-idiotypic antibodies, for affinity		
CC purification of zsig46, to screen expression libraries, to neutralise		
CC zsig46 activity, and to deliver toxins, radioisotopes etc. for		
CC therapeutic or diagnostic purposes. Agonists of the product can be used		

to promote growth, differentiation and proliferation of specific cell types, e.g. for treating (extra)thyroid diseases or as additive to cell cultures

Sequence 346 AA;

Query Match 95.5%; Score 147; DB 2; Length 346;

Best Local Similarity 96.4%; Pred. No. 2.9e-12;

Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MRRGAGARGASWCALALMLAVPG 28  
1 MRRGAGARGASWCALALMLAVPG 28

ESUT 4

AU29418

AU29418 standard; protein; 192 AA.

AU29418;

18-DEC-2001 (first entry)

Human G protein-coupled receptor (GPCR) polypeptide #39.

Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia; neurological disorder; metabolic disorder; cancer; rheumatoid arthritis; thyroid disorder; neurodegenerative disorder; cardiovascular disorder; renal failure; autoimmune disorder; hyperproliferative disorder; HIV; human immunodeficiency virus; viral infection; neuroprotective; immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant; anorectic; gene therapy.

Homo sapiens.

WO200168858-A2.

20-SEP-2001.

16-MAR-2001; 2001WO-US008456.

16-MAR-2000; 2000US-0187783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

29-MAR-2000; 2000US-0192155P.

29-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192833P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192945P.

(PHAA) PHARMACIA & UPJOHN CO.

Vogeli G;

WPI; 2001-607458/69.

N-PSDB; AAS46857.

Nucleic acid encoding G-protein coupled receptors, useful for the prevention, diagnosis and treatment of mental disorders.

Claim 31; Page 73; 274pp; English.

Sequences AU29380-AU29509 represent human G protein-coupled receptor (GPCR) polypeptides of the invention. The proteins and the DNA sequences encoding them can be used to identify compounds which bind to GPCR polypeptides and in screening for compounds that modulate GPCR activity.

By screening a human subject for the presence of mutations in GPCR DNA, a GPCR-related disorder or a genetic predisposition can be diagnosed. The sequences can also be used for treatment and prevention of mental

disorders such as schizophrenia, neurological disorders such as manic depression, metabolic disorders such as obesity, cancer, rheumatoid arthritis, thyroid disorders such as myxoedema, neurodegenerative disorders such as Parkinson's disease, cardiovascular disorders such as atherosclerosis, renal failure, autoimmune disorders, hyperproliferative disorders such as psoriasis and viral infections such as those caused by HIV

Sequence 192 AA;

Query Match 39.3%; Score 60.5; DB 4; Length 192;

Best Local Similarity 39.3%; Pred. No. 1.8;

Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

4 GAGARGASWCW-----ALALW 22  
35 GKAVRGRCPCWPCQPALVITLWV 62

RESULT 5

ABG60706

ABG60706 standard; protein; 192 AA.

ABG60706;

13-AUG-2002 (first entry)

Novel G protein coupled receptor (nGPCR-x) #39.

G protein coupled receptor; nGPCR-x; immune response; thyroid disorder; mental disorder; chylotoxocosis; myxoedema; inflammatory condition; Crohn's disease; cell differentiation; homeostasis; rheumatoid arthritis; renal failure; autoimmune disorder; movement disorder; CNS disorder; viral infection; human immunodeficiency virus; HIV; metabolic disorder; cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy; poriferative disease; cancer; psoriasis; lung cancer; hormonal disorder; sexual dysfunction.

Homo sapiens.

US2002058306-A1.

16-MAY-2002.

16-MAR-2001; 2001US-00811284.

16-MAR-2000; 2000US-0189783P.

16-MAR-2000; 2000US-0189907P.

16-MAR-2000; 2000US-0189917P.

16-MAR-2000; 2000US-0189918P.

16-MAR-2000; 2000US-0189960P.

29-MAR-2000; 2000US-0192155P.

29-MAR-2000; 2000US-0192234P.

29-MAR-2000; 2000US-0192830P.

29-MAR-2000; 2000US-0192916P.

29-MAR-2000; 2000US-0192923P.

29-MAR-2000; 2000US-0192933P.

29-MAR-2000; 2000US-0192945P.

(VOGE/) VOGELI G.

Vogeli G;

WPI; 2002-434856/46.

N-PSDB; ABR61635.

New isolated nucleic acid encoding a G protein coupled receptor for producing the receptor which can induce an immune response in a mammal.

Claim 27; Page 44; 216pp; English.

The invention describes an isolated nucleic acid (1) comprising a

sequence encoding a portion of a G protein coupled receptor (ngPCR-x).  
(1) is used to produce a recombinant ngPCR-x polypeptide. A polypeptide encoded by (1) is used to induce an immune response in a mammal. ngPCR-x is used to identify a compound that binds to it and/or modulates its activity. (1) is used to identify animal homologues of ngPCR-x. (1) can be used to diagnose a human subject as having a brain or genetic predisposition disorder, such as a mental disorder. (1) is used to screen for an ngPCR-x related disorder including thyroid disorders (e.g. thyrotoxicosis, myxedema), renal failure, inflammatory conditions (e.g. Crohn's disease), diseases related to cell differentiation and homeostasis, rheumatoid arthritis, autoimmune disorders, movement disorders, CNS disorders, viral infections (e.g. diabetes, obesity, virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity, anorexia, cardiomyopathies), proliferative diseases and cancers (e.g. psoriasis, lung cancer), hormonal disorders, sexual dysfunction and hereditary mental disorders in a human patient. A host cell comprising (1) is used to screen for a modulator of ngPCR-x activity. ngPCR-x is used to identify compounds that can treat mental disorders. The polypeptide encoded by (1) is used to purify a G protein from a sample. This is the amino acid sequence of a novel G protein coupled receptor (ngPCR-x) protein described in the invention

Q: Sequence 192 AA:

Query Match 39.3%; Score 60.5; DB 5; Length 192;  
Best Local Similarity 39.3%; Pred. No. 1.8;  
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

Y 4 GAGAGRGASWCW-----ALALIM 22  
b 35 GKAAVGRPCWCPCPCPALVSIILAW 62

RESULT 6  
AAU14162 standard; protein; 127 AA.

AAU14162;

24-OCT-2001 (first entry)

Human novel protein #33.

Human; novel protein; Anti-anemic; osteopathic; anti-inflammatory;  
immunomodulatory; cytosolic; neuroprotective; valnerary; nocotropic;  
anticonvulsant; antiarthritic; cerebroprotective; antitumoral; antiviral;  
antibacterial; immunologic; dermatological; haemostatic; antiasthmatic;  
thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
tissue regeneration; immune disorder.

Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002623.

25-JAN-2000; 2000US-00491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

N-PSDB; AAS22467.

Isolated polypeptides useful for treating anti-inflammatory diseases,  
nervous system disorders, and for regenerating bone and cartilage.

Example 4; Page 543-544; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. CC Polynucleotides of the invention are used as probes and primers, for CC sequencing, for chromosome or gene mapping, in the production of CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene CC therapy. Polypeptides of the invention can be used to target drugs to a CC tumour, in assays to determine biological activity, to raise CC antibodies/elicite an immune response, to determine quantitative protein CC levels, as tissue markers, and to isolate receptors or ligands. CC polypeptides of the invention may also be useful in treating platelet CC disorders, stem cell disorders, regenerating bone, cartilage, tendon, CC ligament and/or nerve tissue, wound healing, treating burns, promoting CC the proliferation, differentiation and survival of stem cells, as a CC contraceptive, treating osteoporosis and osteoarthritis, anaemia, CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft- CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory CC diseases, nervous system disorders, and infection. The present sequence CC represents a protein of the invention

Q: Sequence 127 AA:

Query Match 38.6%; Score 59.5; DB 4; Length 127;  
Best Local Similarity 40.0%; Pred. No. 1.6;  
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

QY 3 RGAGARGR-----ASWCALALD--FLAVVP 28  
b 25 RRGRTWRVMTKIAQMDLGLIAGSTVALITG 59

RESULT 7  
ABU60973 standard; protein; 188 AA.

ABU60973;

08-MAY-2003 (first entry)

Lung specific protein (LSP) #76.

Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;  
cancer monitoring; cancer staging; cancer imaging; lung cancer;  
non-cancerous diseases of the lung; transgenic animal.

Homo sapiens.

WO200268633-A2.

06-SEP-2002.

21-NOV-2001; 2001WO-US043612.

22-NOV-2000; 2000US-0252500P.

(DIAD-) DIADEXUS INC.

Macina RA, Recipon H, Chen S, Sun Y, Liu C;

WPI; 2002-713376/77.

Isolated human nucleic acid molecule and polypeptide, useful for  
identifying, diagnosing, monitoring, staging, imaging and treating lung  
cancer and non-cancerous diseases of the lung.

Claim 11; Page 359-360; 389pp; English.

C The invention describes an isolated human nucleic acid (I) encoding any  
C of 120 10-1533 residue amino acid sequences (S1), given in the  
C specification, comprising any of 164 179-12421 base pair sequences (S2),  
C given in the specification. The methods and compositions of the present  
C invention are useful for identifying, diagnosing, monitoring, staging,  
C imaging and treating lung cancer and non-cancerous diseases of the lung.  
C They are also used for identifying lung tissue, monitoring and  
C identifying and/or designing antagonists of the polypeptide of the  
C invention, gene therapy, production of transgenic animals and production  
C of engineered lung tissue for treatment and research. This is the amino  
C acid sequence of a lung specific nucleic acid

Q Sequence 188 AA;

Query Match 37.7%; Score 58; DB 5; Length 188;  
Best Local Similarity 38.7%; Pred. No. 3.9;  
Matches 12; Conservative 1; Mismatches 12; Indels 6; Gaps 1;

Y 4 GAGAGRGRASWCM-----ALALMLAVPG 28  
b 19 GGGGAGAGSWWGMGCGAGALMVAVVG 49

RESULT 8  
BP76682

D ABP76682 standard; protein; 19938 AA.

C ABP76682;

T 26-FEB-2003 (first entry)

X Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.

W Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;

W medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.

X Streptomyces viridochromogenes.

N WO200268436-A1.

D 06-SEP-2002.

X 24-APR-2001; 2001WO-EP009815.

R 25-FEB-2001; 2001DE-01009166.

A (COMB-) COMBINATURE BIOPHARM AG.

I Weitzner G, Muehlenweg A, Trefzer A, Bechtold A;

R MPI; 2003-018650/01.

R N-PSDB; AB237516.

T New avilamycin derivatives, useful for treatment of infections, and  
nucleic acid encoding avilamycin synthesis enzymes.

S Example 1; Page 68-301; 319pp; German.

X The invention relates to avilamycin derivatives (I) with antibacterial,  
C virucide, protozoacide and fungicide activity. (I) are useful for  
C treatment of infections (bacterial, viral, protozoal or fungal), in human  
C or veterinary medicine, particularly where caused by Staphylococcus  
C aureus. (I) are more hydrophilic than known avilamycins. The present  
C sequence is that of an avilamycin synthesis enzyme from the Streptomyces  
C viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-  
C AB237516)

Q Sequence 19938 AA;

Query Match 37.3%; Score 57.5; DB 6; Length 19938;  
Best Local Similarity 41.4%; Pred. No. 5.3e+02;  
Matches 12; Conservative 3; Mismatches 11; Indels 3; Gaps 1;

OY 2 RGAG---GAAGRASWCMALALMLAVVP 27  
Db 2905 RRAATPAGAGRVCMGCPVSGWLCAPF 2933

RESULT 9  
ABU05547

ID ABU05547 standard; protein; 214 AA.

AC ABU05547;

DT 08-APR-2003 (first entry)

DE M. tuberculosis and M. leprae marker protein #198.

XX Mycobacterioses; survival; virulence; protective antigen; vaccine;

KW Mycobacterial disease; tuberculosis; leprosy.

OS Mycobacterium tuberculosis.

EN WO200274903-A2.

PD 26-SEP-2002.

PF 22-FEB-2002; 2002MO-IB001973.

PR 22-FEB-2001; 2001US-0270123P.

PA (INSP ) INST PASTEUR.

PI Cole S;

DR MPI; 2002-759885/82.

PT Identifying and selecting genes for survival or virulence of mycobacteria

PT by a comparative genomic analysis of the sequences of Mycobacterium

PS Claim 17; Page 376-377; 874pp; English.

XX This invention relates to a novel method for identifying essential genes  
CC for survival or virulence of mycobacteria species. The method comprises  
CC aligning the genomic sequence of a first mycobacterium species on a  
CC genomic sequence of a second mycobacterium species and selecting a  
CC polynucleotide sequence that is highly conserved in both genomes with no  
CC counterparts in other bacterial genomic sequences and that corresponds to  
CC an essential gene for the survival or virulence of mycobacterium species.  
CC The method of the invention is useful for detecting M. tuberculosis or M.  
CC leprae infection. The method reduces the number of potential new targets  
CC and protective antigens for new drugs and vaccine compositions to treat  
CC and prevent mycobacterial diseases, particularly tuberculosis and  
CC leprosy. The present sequence represents a marker protein from  
CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the  
CC method of the invention

SQ Sequence 214 AA;

Query Match 36.4%; Score 56; DB 5; Length 214;  
Best Local Similarity 41.7%; Pred. No. 8.4;  
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

OY 3 RGAGAGRASWCMALALMLAVV 26  
Db 175 RTSATARKGYQWRWSIATLMFTTV 198

RESULT 10

ID ABU05190 standard; protein; 403 AA.

AC ABU05190;

XX



20-MAY-2003 (first entry)  
Human NOV105a protein.  
NOVX; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;  
hypocensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;  
human.  
Homo sapiens.  
MO200272757-A2.  
19-SEP-2002.  
08-MAR-2002; 2002MO-US006908.  
08-MAR-2001; 2001US-0274101P.  
08-MAR-2001; 2001US-0274194P.  
08-MAR-2001; 2001US-0274281P.  
08-MAR-2001; 2001US-0274322P.  
09-MAR-2001; 2001US-0274849P.  
12-MAR-2001; 2001US-0275235P.  
13-MAR-2001; 2001US-0275578P.  
13-MAR-2001; 2001US-0275579P.  
13-MAR-2001; 2001US-0275601P.  
14-MAR-2001; 2001US-0276000P.  
16-MAR-2001; 2001US-0276776P.  
19-MAR-2001; 2001US-0276994P.  
20-MAR-2001; 2001US-0277239P.  
20-MAR-2001; 2001US-0277321P.  
20-MAR-2001; 2001US-0277327P.  
21-MAR-2001; 2001US-0277919P.  
22-MAR-2001; 2001US-0278332P.  
23-MAR-2001; 2001US-0278152P.  
26-MAR-2001; 2001US-0278894P.  
27-MAR-2001; 2001US-0279036P.  
28-MAR-2001; 2001US-0279344P.  
30-MAR-2001; 2001US-0279388P.  
30-MAR-2001; 2001US-0279959P.  
30-MAR-2001; 2001US-0280233P.  
02-APR-2001; 2001US-0280822P.  
02-APR-2001; 2001US-0280900P.  
02-APR-2001; 2001US-0281194P.  
04-APR-2001; 2001US-0281194P.  
13-APR-2001; 2001US-0283675P.  
30-APR-2001; 2001US-0287424P.  
02-MAY-2001; 2001US-0288066P.  
03-MAY-2001; 2001US-0288342P.  
03-MAY-2001; 2001US-0288528P.  
15-MAY-2001; 2001US-0291150P.  
16-MAY-2001; 2001US-0291099P.  
16-MAY-2001; 2001US-0291240P.  
30-MAY-2001; 2001US-0294485P.  
31-MAY-2001; 2001US-0294889P.  
31-MAY-2001; 2001US-0294899P.  
18-JUN-2001; 2001US-0299027P.  
19-JUN-2001; 2001US-0299303P.  
19-JUN-2001; 2001US-0299310P.  
10-JUL-2001; 2001US-0304354P.  
31-JUL-2001; 2001US-0309198P.  
16-AUG-2001; 2001US-0312903P.  
10-SEP-2001; 2001US-0318462P.  
12-SEP-2001; 2001US-0318770P.  
27-SEP-2001; 2001US-0325430P.  
27-SEP-2001; 2001US-0325681P.  
18-OCT-2001; 2001US-0330380P.  
31-OCT-2001; 2001US-0335301P.  
14-NOV-2001; 2001US-0332172P.  
14-NOV-2001; 2001US-0332271P.  
14-NOV-2001; 2001US-0332272P.  
14-NOV-2001; 2001US-0333184P.  
14-NOV-2001; 2001US-0333272P.

21-NOV-2001; 2001US-0332094P.  
03-DEC-2001; 2001US-0337426P.  
03-DEC-2001; 2001US-0338092P.  
04-DEC-2001; 2001US-0337185P.  
03-JAN-2002; 2002US-0345705P.  
07-MAR-2002; 2002US-00092900.  
(CURA-) CURAGEN CORP.  
Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CE, Li L;  
Zerhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R;  
Paturajan M, Gangoli E, Vernet CM, Guo X, Tchernov V;  
Fernandes ER, Casman SJ, Malyankar UM, Gerlach V, Liu Y, Anderson D;  
Spaderna SK, Catterton E, Burgess C, Leite M, Zhong H, Alsbrook JP;  
Lepley DM, Rieger DK;  
WPI; 2002-723332/78.  
N-PSDB; ABX97157.  
NOVX polypeptides and polynucleotides, useful for preventing or treating  
a disorder associated with aberrant NOVX expression or activity e.g.,  
cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial  
asthma.  
Claim 1; Page 410; 1103pp; English.  
This invention describes novel human NOVX polypeptides which have  
cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and hypotensive  
activity. Pharmaceutical compositions comprising the NOVX proteins or  
nucleic acid molecules or NOVX antibodies are useful for preventing or  
treating a disorder associated with aberrant NOVX expression or activity  
e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial  
asthma. The products of the invention can be used for gene therapy or in  
a vaccine. AB065041-AB065218 represent the NOVX polypeptides encoded by  
ABX97008-ABX97185  
SQ Sequence 403 AA;  
Query Match 35.7%; Score 55; DB 5; Length 403;  
Best Local Similarity 46.2%; Pred. No. 22;  
Matches 12; Conservative 1; Mismatches 13; Indels 0; Gaps 0;  
QY 2 RRGAGARGRASCWALALMLAVVP 27  
Db 118 RRGSLRGAAGLAWRLFLAVTLV 143  
RESULT 11  
ABBI6658  
ID ABBI6658 standard; protein; 78 AA.  
AC ABBI6658;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polypeptide SEQ ID NO 5315.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
antiParkinsonian; antisticking; antianaemic; antiarthritic; cancer;  
antiheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
antiallergic; antidiabetic; anticancer; anticonvulsant; antifungal;  
antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
neurological disease; infection; nephrotropic; gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN MO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001334.  
XX





polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 211 AA;

Query Match 34.4%; Score 53; DB 4; Length 211;  
Best Local Similarity 34.6%; Pred. No. 22;  
Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

1 MRGAGARGASMCWALALMLAVV 26  
157 MRSVEGFSRSLFECYILSLYVVSAL 182

ABG07764

ABG36249 standard; protein; 346 AA.

ABG36249;

10-DEC-2002 (first entry)

Maize peroxidase protein Zm-POX07.

Peroxidase-like activity; expression cassette; plant defence response; plant disease resistance; plant stalk strength; plant breeding; maize; plant; peroxidase.

Zea mays.

WO200270723-A2.

12-SEP-2002.

17-JAN-2002; 2002WO-US001454.

18-JAN-2001; 2001US-0262595P.  
16-JAN-2002; 2002US-00047825.

(PION-) PIONEER HI-BRED INT INC.

Duvick JP, Maddox JR, Navarro Acevedo PA, Simmons CR;  
WPI; 2002-723267/78.  
N-PSDB; ABS74468.

New maize peroxidase nucleic acid molecule and polypeptide used in expression cassettes for modulating plant defense response, increasing plant disease resistance and increasing plant stalk strength.

Claim 1; Page 87-88; 119pp; English.

The present invention relates to a new nucleic acid molecule that encodes a polypeptide with peroxidase-like activity. The nucleic acid of the invention and all methods are useful for expression cassettes for modulating plant defence response, increasing plant disease resistance and increasing plant stalk strength. They can also be used in methods of selecting or breeding for plants with increased disease resistance. The present amino acid sequence represents a maize peroxidase protein of the invention

Sequence 346 AA;

Query Match 34.4%; Score 53; DB 5; Length 346;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 12; Conservative 2; Mismatches 2; Indels 8; Gaps 2;

3 RGAGARGASMCWALALMLAVV 26  
15 RGDGA---AAWCA---WVAVV 30

ABG07764

ABG07764 standard; protein; 600 AA.

ABG07764;

13-FEB-2002 (first entry)

Novel human diagnostic protein #7755.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.  
23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Dmanac RT, Liu C, Tang YT;  
WPI; 2001-639362/73.  
N-PSDB; AAS71951.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 38123; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 600 AA;

Query Match 34.4%; Score 53; DB 4; Length 600;  
Best Local Similarity 66.7%; Pred. No. 63;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

10 GRASMCWALAL 21

db 321 GAASWCMVSA 332

Search completed: June 8, 2004, 14:03:56  
Job time : 8.42778 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: June 8, 2004, 14:00:33 ; Search time 2.17778 Seconds  
(without alignments)  
663.762 Million cell updates/sec

file: US-10-010-050A-2\_COPY\_1\_28

effect score: 154

sequence: 1 MRRGAGARGRASGCMALALMLAVYFG 28

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

atched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	34.4	74	4	US-09-621-976-6090
2	52.5	34.1	575	1	US-07-683-957B-2
3	52	33.8	170	4	US-09-252-991A-16705
4	52	33.8	217	4	US-09-252-991A-19862
5	51	33.1	85	4	US-09-431-705-22
6	51	33.1	88	4	US-09-205-258-274
7	51	33.1	1310	4	US-09-170-496D-290
8	51	33.1	1310	4	US-09-364-425B-55
9	50.5	32.6	209	4	US-08-311-731A-18
10	50	32.5	127	4	US-09-252-991A-26303
11	50	32.5	286	4	US-09-252-991A-27451
12	49.5	32.1	92	4	US-09-247-155-98
13	49.5	32.1	196	4	US-09-252-991A-31079
14	49	31.8	718	4	US-09-252-991A-25696
15	49	31.8	1498	4	US-09-252-991A-31234
16	48.5	31.5	251	4	US-09-252-991A-27585
17	48.5	31.5	534	4	US-09-252-991A-17265
18	48	31.2	35	4	US-09-148-545-204
19	48	31.2	170	4	US-09-252-991A-26306
20	48	31.2	215	3	US-09-220-528-104
21	48	31.2	222	4	US-09-148-545-135
22	48	31.2	229	4	US-09-252-991A-22195
23	48	31.2	246	4	US-09-252-991A-25102
24	48	31.2	337	4	US-09-252-991A-19602
25	47	30.5	102	4	US-09-252-991A-23296
26	47	30.5	111	4	US-09-107-532A-4380
27	47	30.5	172	4	US-09-252-991A-28284

28	47	30.5	177	4	US-09-252-991A-24167	Sequence 24167, A
29	47	30.5	250	4	US-09-252-991A-24075	Sequence 24075, A
30	47	30.5	439	4	US-09-489-039A-13336	Sequence 13336, A
31	46.5	30.2	308	4	US-09-252-991A-18097	Sequence 18097, A
32	46.5	30.2	418	4	US-09-252-991A-17598	Sequence 17598, A
33	46	29.9	191	2	US-08-469-412A-4	Sequence 4, Appl1
34	46	29.9	191	3	US-09-021-715-4	Sequence 28415, A
35	46	29.9	193	4	US-09-252-991A-28415	Sequence 28054, A
36	46	29.9	200	4	US-09-252-991A-28054	Sequence 176, App
37	46	29.9	248	4	US-09-673-395A-176	Sequence 15, Appl1
38	46	29.9	335	4	US-09-570-856B-15	Sequence 25031, A
39	46	29.9	478	4	US-09-252-991A-25031	Sequence 22787, A
40	46	29.9	560	4	US-09-252-991A-26787	Sequence 23246, A
41	46	29.9	576	4	US-09-252-991A-23246	Sequence 16906, A
42	45.5	29.5	263	4	US-09-252-991A-16906	Sequence 11, Appl
43	45.5	29.5	270	4	US-09-679-686B-10	Sequence 311, App
44	45.5	29.5	272	4	US-09-489-847-311	Sequence 232, App
45	45.5	29.5	273	4	US-09-489-847-232	

#### ALIGNMENTS

##### RESULT 1

US-09-621-976-6090

Sequence 6090, Application US/09621976

Patent No. 663963

GENERAL INFORMATION:

APPLICANT: Dumas Mline Edwards, J.B.

APPLICANT: Jobert, S.

APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: ESTs and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2

CURRENT APPLICATION NUMBER: US/09/621.976

NUMBER OF FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent .pm

SEQ ID NO 6090

LENGTH: 74

TYPE: PRT

ORGANISM: Homo sapiens

US-09-621-976-6090

Query Match

Best Local Similarity 39.3%, Pred. No. 1.6%

Matches 11; Conservative 3; Mismatches 10; Indels 4; Gaps 1;

DB 4 GAGAGGAA-----SQCWALALMLAVVP 27

13 GLGAPGRGEPGQGMCMARSSIFRCYFP 40

RESULT 2

US-07-683-957B-2

Sequence 2, Application US/07683957B

Patent No. 5310880

GENERAL INFORMATION:

APPLICANT: Donahoe, Patricia K.

APPLICANT: Raglin, Richard C.

APPLICANT: MacLaughlin, David T.

TITLE OF INVENTION: Purification of M llerian Inhibiting

TITLE OF INVENTION: Substance

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible



```
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 274
LENGTH: 88
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

; NAME/KEY: SITE
; LOCATION: (53)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals stop translation
US-09-205-258-274

Query Match          33.1%; Score 51; DB 4; Length 88;
Best Local Similarity 56.5%; Pred. No. 3.6;
Matches 13; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY      1  KRGA--GARGRASCWALAL 21
      | : || || || || || ||
Db      44  MAKGRPGAKRGKRWGLAYTLL 66

RESULT 7
US-09-170-496D-290
; Sequence 290, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-(
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 290
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-290

Query Match          33.1%; Score 51; DB 4; Length 1310;
Best Local Similarity 52.9%; Pred. No. 59;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      11  RASMCALALWLAIVP 27
      : |||| | : || |
Db      907  QASWCHARLWVHSGAP 923

RESULT 8
US-09-364-425B-55
; Sequence 55, Application US/09364425B
; Patent No. 6653086
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lin, I-lin
; APPLICANT: Lowitz, Kevin P.
; APPLICANT: Chen, Kuoping
; TITLE OF INVENTION: Endogenous, Constitutively Activated G Protein-Coupled Orphan Recei
; FILE REFERENCE: AREN0047
; CURRENT APPLICATION NUMBER: US/09/364,425B
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/094,879
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: 60/106,300
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/110,906
; PRIOR FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 60/121,851
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
```



SEQ ID NO 55  
LENGTH: 1310  
TYPE: PRT  
ORGANISM: Homo sapiens  
S-09-364-425B-55

Query Match 33.1%; Score 51; DB 4; Length 1310;  
Best Local Similarity 52.9%; Pred. No. 59;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Y 11 RASWCWALMLAVVP 27  
b 907 QASWCHARLWVHSA 923

RESULT 9  
S-08-311-731A-18  
Sequence 18, Application US/08311731A  
Patent No. 6583266  
GENERAL INFORMATION:

APPLICANT: SMITH, DOUGLAS  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 411  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 209 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIUM TUBERCULOSIS  
S-08-311-731A-18

Query Match 32.8%; Score 50.5; DB 4; Length 209;  
Best Local Similarity 38.7%; Pred. No. 10;  
Matches 12; Conservative 2; Mismatches 12; Indels 5; Gaps 1;

Y 1 MRGAGA-----ARGASCKWALMLAVV 26  
b 163 MRENKSGQERVGAILLYQWRMSIATIMPTV 193

RESULT 10  
S-09-252-991A-26303  
Sequence 26303, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26303  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
S-09-252-991A-26303

Query Match 32.5%; Score 50; DB 4; Length 127;  
Best Local Similarity 47.1%; Pred. No. 7.2;  
Matches 8; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 12 ASWCWALMLAVVP 28  
Db 33 ATWCWTLAFTW---Pg 45

RESULT 11  
US-09-252-991A-27451  
Sequence 27451, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27451  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
S-09-252-991A-27451

Query Match 32.5%; Score 50; DB 4; Length 286;  
Best Local Similarity 57.9%; Pred. No. 17;  
Matches 11; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 3 RGAGARG---RASWCWA 17  
Db 35 RGAGARGDLPFRHGWMA 53

RESULT 12  
US-09-247-155-98  
Sequence 98, Application US/09247155A  
Patent No. 6312922  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
APPLICANT: Ducleret, Aymeric  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: Complementary DNAs  
FILE REFERENCE: GENSIT.021A  
CURRENT APPLICATION NUMBER: US/09/247,155A  
PRIOR FILING DATE: 1999-02-09  
EARLIER APPLICATION NUMBER: 60/074,121  
EARLIER FILING DATE: 1998-02-09  
EARLIER APPLICATION NUMBER: 60/081,563



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on on: June 8, 2004, 14:04:04 ; Search time 5.75556 Seconds

(without alignments)  
1368.675 Million cell updates/sec

title: US-10-010-050A-2\_COPY\_1\_28  
project score: 154  
sequence: 1 MRGAGARGASWCMALMLAVPG 28

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 1155919 seqs, 281338677 residues

total number of hits satisfying chosen parameters: 1155919

minimum DB seq length: 0  
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	346	9	US-09-122-383-2
2	154	100.0	346	12	US-10-653-595-126
3	154	100.0	346	13	US-10-010-050A-2
4	154	100.0	347	12	US-09-397-945-126
5	60.5	39.3	192	9	US-09-811-284-167
6	59.5	38.6	127	15	US-10-291-265-269
7	59	38.3	588	14	US-10-156-761-11538
8	58	37.7	188	9	US-09-989-920-240
9	57	37.0	19608	15	US-10-084-846A-8
10	56	36.4	214	14	US-10-080-170-198
11	55	35.7	403	12	US-10-092-900A-300
12	53	34.4	160	14	US-10-156-761-10933
13	53	34.4	346	12	US-10-047-825-10
14	53	34.4	368	12	US-10-425-114-55731
15	52.5	34.1	199	12	US-10-221-278-727

16	52.5	34.1	199	15	US-10-291-172-727	Sequence 727, App
17	52.5	34.1	333	9	US-09-738-626-4771	Sequence 4771, Ap
18	52.5	34.1	360	12	US-10-282-122A-50396	Sequence 50396, A
19	52.5	34.1	464	12	US-10-424-599-207280	Sequence 207280,
20	52	33.8	82	12	US-10-424-599-167251	Sequence 167251,
21	52	33.8	560	10	US-09-863-776-69	Sequence 69, Appl
22	52	33.8	599	15	US-10-104-047-2306	Sequence 2306, Ap
23	52	33.8	653	10	US-09-863-776-28	Sequence 28, Appl
24	52	33.8	658	9	US-09-759-056-5	Sequence 5, Appl
25	52	33.8	658	9	US-09-901-812-5	Sequence 5, Appl
26	52	33.8	658	10	US-09-863-776-68	Sequence 68, Appl
27	52	33.8	662	10	US-09-863-776-30	Sequence 30, Appl
28	52	33.8	667	9	US-09-759-056-2	Sequence 2, Appl
29	52	33.8	667	9	US-09-901-812-2	Sequence 2, Appl
30	52	33.8	667	10	US-09-863-776-32	Sequence 32, Appl
31	52	33.8	667	10	US-09-863-776-67	Sequence 67, Appl
32	52	33.8	667	12	US-10-219-535-80	Sequence 80, Appl
33	52	33.8	667	12	US-10-232-230-80	Sequence 80, Appl
34	52	33.8	667	12	US-10-232-224-80	Sequence 80, Appl
35	52	33.8	667	14	US-10-237-884-80	Sequence 80, Appl
36	52	33.8	667	14	US-10-230-163-80	Sequence 80, Appl
37	52	33.8	667	14	US-10-230-338-80	Sequence 80, Appl
38	52	33.8	667	14	US-10-218-631-80	Sequence 80, Appl
39	52	33.8	667	14	US-10-230-414-80	Sequence 80, Appl
40	52	33.8	667	14	US-10-216-159A-80	Sequence 80, Appl
41	52	33.8	667	14	US-10-218-849-80	Sequence 80, Appl
42	52	33.8	667	14	US-10-227-873-80	Sequence 80, Appl
43	52	33.8	667	14	US-10-227-883-80	Sequence 80, Appl
44	52	33.8	667	14	US-10-219-076-80	Sequence 80, Appl
45	52	33.8	667	14	US-10-230-434-80	Sequence 80, Appl

## ALIGNMENTS

RESULT 1  
US-09-122-383-2  
Sequence 2, Application US/09122383A  
Patent No. US20020042093A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
FILE OF INVENTION: CHROMOSOME 13  
FILE REFERENCE: 97-38  
CURRENT APPLICATION NUMBER: US/09/122,383A  
CURRENT FILING DATE: 1998-07-24  
EARLIER APPLICATION NUMBER: 60/053,613  
EARLIER FILING DATE: 1997-07-24  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-122-383-2

Query Match 100.0%; Score 154; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 6e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;

QY 1 MRGAGARGASWCMALMLAVPG 28  
|||||  
Db 1 MRGAGARGASWCMALMLAVPG 28

RESULT 2  
US-10-653-595-126  
Sequence 126, Application US/10653595  
Publication No. US20040048304A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.  
TITLE OF INVENTION: 95 Human secreted proteins

```
FILE REFERENCE: P2027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-653-595-126

Query Match      100.0%; Score 154; DB 12; Length 346;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28

RESULT 3
US-10-010-050A-2
; Sequence 2, Application US/10010050A
; Publication No. US20020173624A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010,050A
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 09/122,383
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053,613
; PRIOR FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Homo sapien
; US-10-010-050A-2
```

```
Query Match      100.0%; Score 154; DB 13; Length 346;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28

RESULT 4
US-09-397-945-126
; Sequence 126, Application US/09397945
; Publication No. US20030065139A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 95 Human secreted proteins
; FILE REFERENCE: P2027P1
; CURRENT APPLICATION NUMBER: US/09/397,945
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,581
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,577
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,563
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,313
; PRIOR FILING DATE: 1998-04-01
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (242)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (246)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (347)
; OTHER INFORMATION: Xaa equals stop translation
US-09-397-945-126

Query Match      100.0%; Score 154; DB 12; Length 347;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 MRGGAAGRGRASWCWALALMLAVPG 28
Db 1 MRGGAAGRGRASWCWALALMLAVPG 28

RESULT 5
```

```

S-09-811-284-167
Sequence 167, Application US/09811284
Patent No. US20020058306A1
GENERAL INFORMATION:
APPLICANT: Vogeli, Gabriel
TITLE OF INVENTION: No. US20020058306A1 G Protein-Coupled Receptors
FILE REFERENCE: 00167US1
CURRENT APPLICATION NUMBER: US/09/811,284
CURRENT FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/189,783
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,907
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,918
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,960
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,917
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/192,945
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,916
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,923
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,933
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,830
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,234
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/192,155
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 60/192,935
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 258
SOFTWARE: PatentIn version 3.0
SEQ ID NO 167
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
S-09-811-284-167

Query Match          39.3%; Score 60.5; DB 9; Length 192;
Best local similarity 39.3%; Pred. No. 8.3;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1.

      4 GAGARGASGWCW-----ALAILM 22
      |||||
1b 35 GAAVGRGPCWCWPCPALIVSTITALW 62

RESULT 6
IS-10-291-265-269
Sequence 269, Application US/10291265
Publication No. US20030232054A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
APPLICANT: Tang et al
TITLE OF INVENTION: No. US20030232054A1 Nucleic Acids and Polypeptides
FILE REFERENCE: 2172-017 (785)
CURRENT APPLICATION NUMBER: US/10/291,265
CURRENT FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/617,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0

```

```

      / SEQ ID NO 269
      / LENGTH: 127
      / TYPE: PRT
      / ORGANISM: Homo sapiens
      / US-10-291-265-269

Query Match
Best Local Similarity 38.6%; Score 59.5; DB 15; Length 127;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

QY
  3 RGAGARGR-----ASNCWALAIL--WLAVPG 28
    ||| ||| ||| ||| ||| ||| ||| |||
Db 25 RGEETWGRVTMTKLAQWLWGLAIIGSTVALTTG 59

RESULT 7
US-10-156-761-11538
/ Sequence 11538, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMIURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 11538
/ LENGTH: 588
/ TYPE: PRT
/ ORGANISM: Streptomyces avermiltilis
US-10-156-761-11538

Query Match
Best Local Similarity 38.3%; Score 59; DB 14; Length 588;
Matches 13; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

QY
  3 RGAGAR-----GRASWCWALALLWLA 24
    ||| ||| ||| ||| ||| ||| ||| |||
Db 167 RGAGSSRTKLYATKMLTELIGVYGSRYGGDKSWSMAGGLWLA 210

RESULT 8
US-09-989-920-240
/ Sequence 240, Application US/09989920
/ Patent No. US20020172957A1
/ GENERAL INFORMATION:
/ APPLICANT: Macina, Roberto
/ APPLICANT: Recipon, Hervé
/ APPLICANT: Chen, Sei-Yu
/ APPLICANT: Sun, Yongming
/ APPLICANT: Liu, Chenghua
/ TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
/ FILE REFERENCE: DEX-0291
/ CURRENT APPLICATION NUMBER: US/09/989,920
/ CURRENT FILING DATE: 2001-11-21
/ PRIOR APPLICATION NUMBER: 60/252,500
/ PRIOR FILING DATE: 2000-11-22
/ NUMBER OF SEQ ID NOS: 284
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 240
/ LENGTH: 188
/ TYPE: PRT
/ ORGANISM: Homo sapien

```

JS-09-989-920-240

Query Match 37.7%; Score 58; DB 9; Length 188;  
Best Local Similarity 38.7%; Pred. No. 17;  
Matches 12; Conservative 1; Mismatches 12; Indels 6; Gaps 1;

Dy 4 GAGAGRAGSWCM-----ALALMLAVPG 28  
Db 19 GGGGAGAGSWVMWCGCGGAGMLVAVGCG 49

RESULT 9

JS-10-084-846A-8  
; Sequence 8, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLBENWEG, AGNES  
; APPLICANT: TREFFER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084, 846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 8  
; LENGTH: 19608  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-8

Query Match 37.0%; Score 57; DB 15; Length 19608;  
Best Local Similarity 42.9%; Pred. No. 1.2e+03;  
Matches 12; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Dy 2 RRGAG--GAARGASWCMALALMLAVVP 27  
Db 2876 RRAAVTPGARGRVEKCPVSGVWLCAFP 2903

RESULT 10

US-10-080-170-198  
; Sequence 198, Application US/10080170  
; Publication No. US20030129601A1  
; GENERAL INFORMATION:  
; APPLICANT: COLE, S.T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
; FILE REFERENCE: 03495, 0218  
; CURRENT APPLICATION NUMBER: US/10/080, 170  
; CURRENT FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270, 123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 198  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-080-170-198

Query Match 36.4%; Score 56; DB 14; Length 214;  
Best Local Similarity 41.7%; Pred. No. 32;  
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Dy 3 RGAGARGASWCMALALMLAVV 26  
Db 175 RTSAIRGIYQWRNSTALWFLITV 198

RESULT 11

US-10-092-900A-300  
; Sequence 300, Application US/10092900A  
; Publication No. US20040043382A1  
; GENERAL INFORMATION:  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Taupler Jr., Raymond J.  
; APPLICANT: Pena, Carol E.A.  
; APPLICANT: Li, Li  
; APPLICANT: Zernusen, Bryan D.  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Gorman, Linda  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Patlurajan, Meera  
; APPLICANT: Gangoli, Bsha A.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Guo, Xiaojia Sasha  
; APPLICANT: Tcherenev, Velizar T.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Liu, Yi  
; APPLICANT: Anderson, David W.  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Catterton, Elina  
; APPLICANT: Leite, Mario W.  
; APPLICANT: Zhong, Haihong  
; APPLICANT: Alsobrook, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Bugess, Catherine E.  
; TITLE OF INVENTION: No. US20040043382A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-290C  
; CURRENT APPLICATION NUMBER: US/10/092, 900A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: USSN 60/274,322  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/283,675  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: USSN 60/338,092  
; PRIOR FILING DATE: 2001-12-03  
; PRIOR APPLICATION NUMBER: USSN 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/274,191  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/325,681  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: USSN 60/304,354  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: USSN 60/279,995  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: USSN 60/294,899  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: USSN 60/287,424  
; PRIOR FILING DATE: 2001-04-30  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 768  
; SEQ ID NO 300  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-092-900A-300



Query Match 34.1%; Score 52.5; DB 12; Length 199;  
Best Local Similarity 73.3%; Pred. No. 81;  
Matches 11; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

ly 2 RRGGGAARGRASWCW 16  
| | | | | | | | | |  
yb 164 RDGAGRAAGRA-WYW 177

Search completed: June 8, 2004, 14:11:57  
Job time : 6.75556 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - protein search, using sw model

on: June 8, 2004, 13:59:38 ; Search time 2.02222 Seconds  
(without alignments)  
1331.683 Million cell updates/sec

File: US-10-010-050a-2\_COPY\_1\_28  
Effect score: 154  
Sequence: 1 MRGAGARGRACWCMALALIMLAVVPG 28

oring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

set-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	42.9	606	2 C84390	sulfate transport
2	59	38.3	412	2 D95941	conserved hypothet
3	56	36.4	214	2 T44701	probable integral
4	54	35.1	232	2 F82729	conserved hypothet
5	53	34.4	428	2 A12860	conserved hypothet
6	53	34.4	428	2 H97637	hypothetical prote
7	52.5	34.1	575	1 WPROM	multimeric inhibiti
8	52	33.8	641	2 A45054	probable intercell
9	51.5	33.4	686	2 S30075	ferric reductase (
10	51.5	33.4	1146	2 A38587	collagen, cornea-s
11	51	33.1	410	2 E95984	conserved hypothet
12	50.5	32.8	210	2 E70671	hypothetical prote
13	50	32.5	106	2 A29760	hypothetical prote
14	50	32.5	163	2 S73039	hypothetical prote
15	50	32.5	221	2 G83964	hypothetical prote
16	50	32.5	249	2 B84147	ABC transporter (p
17	50	32.5	250	2 AD1197	ABC transporter tr
18	50	32.5	417	2 AF3448	hypothetical membr
19	49.5	32.1	201	2 D81944	probable integral
20	49	31.8	118	2 S52855	hypothetical prote
21	49	31.8	139	2 A71123	hypothetical prote
22	49	31.8	188	2 B82183	ank protein VC158
23	49	31.8	250	2 AB1555	ABC transpore tr
24	49	31.8	310	2 T16233	hypothetical prote
25	49	31.8	317	2 T35010	probable integral
26	49	31.8	755	2 B75346	probable compenac
27	49	31.8	1221	2 B83327	conserved hypothet
28	48.5	31.5	210	2 S36297	T-cell receptor ga
29	48.5	31.5	435	2 A12082	hypothetical prote

30	48.5	31.5	519	2 S75570	apolipoprotein N-a
31	48.5	31.5	688	2 B97152	probable membrane
32	48.5	31.5	840	2 AG0526	penicillin-binding
33	48.5	31.5	854	2 D82324	conserved hypothet
34	48	31.2	159	2 A48426	homeotic protein H
35	48	31.2	221	2 UC7587	hypothetical cell-deriv
36	48	31.2	233	2 B40364	stromal cell-deriv
37	48	31.2	334	2 F75539	branched-chain am
38	48	31.2	365	2 AB2941	hypothetical prote
39	48	31.2	369	2 H98341	probable di-trans,
40	48	31.2	400	2 T35334	probable membrane
41	48	31.2	968	2 T00322	hypothetical prote
42	48	31.2	1245	1 MMSND	nidogen precursor
43	47.5	30.8	519	2 F96517	hypothetical prote
44	47.5	30.8	737	2 H96538	hypothetical prote
45	47.5	30.8	1360	2 JC5839	GTBR-N protein - h

#### ALIGNMENTS

##### RESULT 1

C84390 sulfate transport system permease protein [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84390  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Lechhauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lie A:Title: Genome sequence of Halobacterium species NRC-1  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: C84390  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-606 <STO>  
A:Cross-references: GB:AB004437; NID:g10581803; PIDN:AA620487.1; GSPDB:GN00138  
A:Gene: cyst1

Query Match 42.9%; Score 66; DB 2; Length 606;  
Best Local Similarity 53.8%; Pred. No. 0.31;  
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 3 RCAGARGRACWCMALALIMLAVVPG 28  
DB 508 RALGARGRACWCMALALIMLAVVPG 533

##### RESULT 2

D95941 conserved hypothetical membrane protein, paralogue of Y20848 SMB21292 [imported] - Sinorhizobium meliloti  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: D95941  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmesier, J.; Chair, P.; Vorholter, F.J.; Hernan Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: D95941  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-412 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CA049196.1; PID:g15140681; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chair, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, A.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.

```

>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
>Reference number: A96039; MUID:21368234; PMID:11474104
>Contents: annotation
>Genetics:
>Gene: Smb21292
>Genome: plasmid

Query Match          38.3%; Score 59; DB 2; Length 412;
Best Local Similarity 37.0%; Pred. No. 1.9;
Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

y      7 AAGRASWCWA-----LALLWLAVVP 27
      | : | : | : | : | : | : | : | : |
b      276 ANKGRETWTWLTATVLLFLIAIMWLTVP 302
      | : | : | : | : | : | : | : | : |

RESULT 3
Probable integral membrane protein [imported] - Mycobacterium leprae
>Species: Mycobacterium leprae
>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #ext_change 28-Jul-2000
>Accession: T44701
>Parkhill, J., Barrell, B.G., Rajandream, M.A.
Submitted to the EMBL Data Library, May 1998
>Reference number: Z22830
>Accession: T44701
>Status: preliminary; translated from GB/EMBL/DBJ
>Molecule type: DNA
>Residues: 1-214 <PAR>
>Cross-references: EMBL:AL023635; PIDN:CAP19188.1
>Experimental source: cosmid B1243
>Genetics:
>Note: MCB1243.07
>Superfamily: Mycobacterium tuberculosis hypothetical protein RV2968c

Query Match          36.4%; Score 56; DB 2; Length 214;
Best Local Similarity 41.7%; Pred. No. 2.8;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

y      3 RGAGGASGASWCWALALIMLAVV 26
      | : | | | : | : | : | : | : | : |
b      175 RFSAIRGIYQWRMSIATLFTITV 198
      | : | : | : | : | : | : | : | : |

RESULT 4
Conserved hypothetical protein XP1054 [imported] - Xylella fastidiosa (strain 945c)
>Species: Xylella fastidiosa
>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 15-Sep-2000
>Accession: F82729
>Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
ature 406, 151-157, 2000
>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
>Reference number: A82515; MUID:20365717; PMID:10910347
>Note: for a complete list of authors see reference number A59328 below
>Accession: F82729
>Status: preliminary
>Molecule type: DNA
>Residues: 1-232 <STM>
>Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83864.1; GSPDB:GN001
>Experimental source: strain 945c
>Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carver, H.
de-Nevo, E.; Docena, C.; El-Dorri, H.; Recincant, A.P.; Ferreira, A.J.S.
Submitted to Genbank, June 2000
>Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; France, S.C.; Franco, M.C.; Frohnt
J.D.; Jungueira, M.D.; Kemper, E.D.; Kitejima, J.P.; Krieger, U.E.; Kuramae, E.B.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A.; Authors: Martins, E.M.F.; Matsunuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaasak
e, A.; Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
, Tshachko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

```

```

A:Reference number: A59238
A:Contents: annotation
C:Genetics:
A:Gene: XFI054
C:Superfamily: Streptomyces coelicolor probable integral membrane protein SC6G10.12

Query Match
Best Local Similarity 35.1%; Score 54; DB 2; Length 232;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 AARGASMCWALALMLAVP 26
Db 13 SRNPRTMLAVPLLEFAVL 31

RESULT 5
A:12860
Conserved hypothetical protein Atu2314 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: A12860
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: A12860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL4303.1; PID:g17740794; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2314
A:Map position: circular chromosome

Query Match
Best Local Similarity 34.4%; Score 53; DB 2; Length 428;
Matches 8; Conservative 7; Mismatches 6; Indels 6; Gaps 1;

QY 7 AARGASMCWA-----LALMLAVP 27
Db 284 ARKGKPTWLTALFLIYIMLSTVP 310

RESULT 6
H97637
Hypothetical protein AGR_C_4211 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: H97637
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2333-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97637
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88057.1; PID:g15157481; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4211
A:Map position: circular chromosome

Query Match
Best Local Similarity 34.4%; Score 53; DB 2; Length 428;
Matches 8; Conservative 7; Mismatches 6; Indels 6; Gaps 1;

QY 7 AARGASMCWA-----LALMLAVP 27

```

b |:::| |:::| |  
284 ARKGKPTWLTATLAFIVIMWSTVP 310

# RESULT 7

FBOM

ulterian inhibiting factor precursor - bovine  
;Alternate names: Mullerian inhibiting substance (MIS)

;Species: Bos primigenius taurus (cattle)

;Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #ext\_change 01-Dec-2000

;Accession: A01398; B01398

;Cate: R.L.; Matcaliano, R.U.; Hession, C.; Tizard, R.; Farber, N.M.; Cheung, A.; Ninfu  
n, K.L.; Ragin, R.C.; Mangano, T.F.; MacLaughlin, D.T.; Donahoe, P.K.

;ell 45; 685-698; 1986

;Title: Isolation of the bovine and human genes for Mullerian inhibiting substance and

;Reference number: A90879; PMID:86218082; PMID:3754790

;Accession: A01398

;Molecule type: DNA

;Residues: 1-14 <CA1>

;Experimental source: newborn calf testis, clones cbm15 and p521

;Accession: B01398

;Molecule type: mRNA

;Residues: 15-575 <CA2>

;Comment: This glycoprotein, produced by the Sertoli cells of the testis, causes regres-

tion of Mullerian duct origin. Other roles for this protein in gonadal differentiation

and duct regression and in the adult ovary.

;Comment: This protein is homologous to the beta transforming growth factor, inhibin al-

these sequences. All of these proteins are biologically active as disulfide-linked dimer

;Comment: Although it does not compete with EGF for receptor binding sites, MIS can inh-

;Superfamily: Inhibin

;Keywords: cytochrome; glycoprotein; gonadal differentiation; testis

;1-19/Domain: signal sequence #status predicted <SIG>

;20-24/Domain: propeptide #status predicted <PRO>

;25-575/Product: mullerian inhibiting factor #status predicted <MIF>

;28-344/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 34.1%; Score 52.5; DB 1; Length 575;

Best Local Similarity 44.0%; Pred. No. 19;

Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

# RESULT 8

45054

robable intercellular signal transducer or transmitter Pz-1 - rat

;Species: Rattus norvegicus (Norway rat)

;Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #ext\_change 02-Mar-2001

;Accession: A45054

;Chan, S.D.; Karpf, D.B.; Fowlkes, M.E.; Hooks, M.; Bradley, M.S.; Vuong, V.; Bambino,

. Biol. Chem. 267, 25202-25207, 1992

;Title: Two homologs of the Drosophila polarity gene frizzled (fz) are widely expressed

;Reference number: A45054; PMID:93094228; PMID:1334084

;Accession: A45054

;Status: preliminary; not compared with conceptual translation

;Molecule type: nucleic acid

;Residues: 1-641 <CHA>

;Experimental source: UMR 106 osteosarcoma cell line

;Note: sequence extracted from NCBI backbone (NCBIP:120154)

;Superfamily: fruit fly frizzled protein

Query Match 33.8%; Score 52; DB 2; Length 641;

Best Local Similarity 55.6%; Pred. No. 24;

Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

# RESULT 9

5

AGAGAGRA-SWCA---LALMLAVP 27

38 AGRRPAPASRCWARGLLMLLEAP 64

# S30075

ferric reductase (EC 1.6.99.-) PRE1 - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein L8167.2; protein YLR214W

C:Species: Saccharomyces cerevisiae

C:Date: 28-May-1993 #sequence\_revision 28-May-1993 #ext\_change 17-Mar-2000

C:Accession: S30075; S48565

R:Pancis, A.; Roman, D.G.; Anderson, G.J.; Hinebsch, A.G.; Klausner, R.D.

Proc. Natl. Acad. Sci. U.S.A. 89, 3869-3873, 1992

A:Title: Ferric reductase of Saccharomyces cerevisiae: Molecular characterization, role i

A:Reference number: S30075; PMID:92237270; PMID:1570306

A:Accession: S30075

A:Molecule type: DNA

A:Residues: 1-686 <DAN>

A:Cross-references: EMBL:86908; NID:g171520; PID:AAA34608.1; PID:g171521

R:Pauley, A.

Submitted to the EMBL Data Library, September 1994

A:Description: The sequence of S. cerevisiae cosmid 8167.

A:Reference number: S48545

A:Accession: S48565

A:Molecule type: DNA

A:Residues: 1-686 <PAU>

A:Cross-references: EMBL:U14913; NID:g544497; PID:AA867424.1; PID:g544499; MIPS:YLR214W

C:Genetics:

A:Gene: PRE1

A:Cross-references: SGD:S0004204; MIPS:YLR214W

A:Map position: 12R

C:Superfamily: ferric reductase PRE2

C:Keywords: oxidoreductase; transmembrane protein

F:6-22/Domain: transmembrane #status predicted <TW1>

F:152-168/Domain: transmembrane #status predicted <TW2>

F:216-232/Domain: transmembrane #status predicted <TW3>

F:258-274/Domain: transmembrane #status predicted <TW4>

F:298-314/Domain: transmembrane #status predicted <TW5>

F:329-345/Domain: transmembrane #status predicted <TW6>

F:359-375/Domain: transmembrane #status predicted <TW7>

F:531-547/Domain: transmembrane #status predicted <TW8>

Query Match 33.4%; Score 51.5; DB 2; Length 686;

Best Local Similarity 47.1%; Pred. No. 29;

Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

# RESULT 10

A38587

collagen, cornea-specific - chicken

C:Species: Gallus gallus (chicken)

C:Date: 14-Feb-1992 #sequence\_revision 15-Aug-1997 #ext\_change 15-Sep-2003

C:Accession: S16501; A38587

R:Marchant, J.K.; Hensenmayer, T.F.; Gordon, M.K.

Proc. Natl. Acad. Sci. U.S.A. 89, 1560-1564, 1991

A:Title: CDNA analysis predicts a cornea-specific collagen.

A:Reference number: A38587; PMID:91142213; PMID:1705041

A:Accession: S16501

A:Molecule type: mRNA

A:Residues: 1-1146 <WAB>

A:Cross-references: EMBL:M60172; NID:g211609; PID:AAA48703.1; PID:g211610

A:Accession: A38587

A:Molecule type: mRNA

A:Residues: 1-174 'X', 176-233 'X', 235-344 'X', 346-408 'X', 410-499 'X', 501-876 'X', 878-114

A:Cross-references: GB:M60172

C:Keywords: cornea

# Query Match

Best Local Similarity 33.4%; Score 51.5; DB 2; Length 1146;

Matches 15; Conservative 2; Mismatches 6; Indels 13; Gaps 3;

4 GAGAGAGRA-SWC-----WALAL--WLAAY 26

29 GAGAGAGAPSWCPGSCGCMWKMGLLMLL 64



;Molecule type: DNA  
;Residues: 1-221 <STO>  
;Cross-references: GB:AP001515; GB:BA000004; NID:G10174886; PIDN:BA06238.1; GSPDB:GN00  
;Experimental source: strain C-125  
;Genetics:  
;Gene: BH2519

Query Match 32.5%; Score 50; DB 2; Length 221;  
Best Local Similarity 41.7%; Pred. No. 18;  
Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 3;

Y 4 GAGAGRGRASWCVALALMLAVVP 27  
D 31 GIGALIGAASFVWFLIAWAILFP 54

Search completed: June 8, 2004, 14:08:08  
3D time : 3.02222 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 1.16667 Seconds

(without alignments)  
1249.684 Million cell updates/sec

file: US-10-010-050a-2\_COPY\_1\_28  
affect score: 154  
sequence: 1 MRRGAGAAAGRAAGCAATATLWLAIVPG 28

oring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

cal number of hits satisfying chosen parameters: 141681

imum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	407	1	CLN5 HUMAN
2	56.5	36.7	126	1	CRGB RALSO
3	52.5	34.1	575	1	MIS BOVIN
4	52	33.8	641	1	FZD1 RAT
5	51.5	33.4	686	1	FRE1 YEAST
6	50	32.5	106	1	SH RAT
7	49.5	32.1	92	1	DEP3 HUMAN
8	49	31.8	286	1	CLN8 HUMAN
9	49	31.8	330	1	PEXD CAEEL
10	48.5	31.5	519	1	LNT SYNY3
11	48	31.2	162	1	HXA9 CAVPO
12	48	31.2	221	1	SDPL HUMAN
13	48	31.2	233	1	PIC1 AGRT5
14	48	31.2	1245	1	NIDO MOUSE
15	47	30.5	339	1	YUGB ECOLI
16	47	30.5	343	1	YUGB ECOLI
17	47	30.5	343	1	GAS1 MOUSE
18	47	30.5	345	1	GAS1 HUMAN
19	47	30.5	398	1	GSPD KLEPN
20	46	30.5	508	1	Y202 HUMAN
21	46	29.9	335	1	XVNB STRLI
22	46	29.9	812	1	UBA4 HUMAN
23	45.5	29.5	108	1	YNPA SALTY
24	45.5	29.5	249	1	MSX1 CHICK
25	45.5	29.5	844	1	PBBP ECOLI
26	45.5	29.5	1112	1	CN3B ECOLI
27	45.5	29.5	2832	1	NR2B RHIME
28	45	29.2	359	1	CDL2 MOUSE
29	45	29.2	362	1	CNTR CHICK
30	45	29.2	755	1	LEB2 RHIPU
31	45	29.2	1021	1	CARA MOUSE
32	45	29.2	1041	1	CHS1 CRVNE
33	44.5	28.9	92	1	DPW3 MOUSE

34	44.5	28.9	108	1	YNPA SALTY	O82622 salmoneila
35	44.5	28.9	307	1	RCGM RHOSH	P02853 rhodobacter
36	44.5	28.9	420	1	MLIC XENLA	P49219 xenopus lae
37	44.5	28.9	451	1	NORM DEIRA	Q91444 deimococcus
38	44.5	28.9	482	1	CLIS2 BACSU	P71040 bacillus su
39	44.5	28.9	758	1	CSTA MYCTU	P95095 mycobacteri
40	44.5	28.9	1356	1	HEM1 PODAN	O08088 podospora a
41	44	28.6	332	1	OSTY YEAST	O03723 saccharomyc
42	44	28.6	442	1	TM14 HUMAN	Q14142 homo sapien
43	44	28.6	459	1	SUOX CHICK	P07850 gallus gall
44	44	28.6	462	1	HEMO HUMAN	P02790 homo sapien
45	44	28.6	560	1	FXJB MOUSE	P59997 mus musculus

## ALIGNMENTS

RESULT 1  
CLN5\_HUMAN  
ID CLN5\_HUMAN STANDARD; PRT; 407 AA.  
AC 075503;  
DT 15-DEC-1998 (rel. 37, Created)  
DT 15-DEC-1998 (rel. 37, Last sequence update)  
DT 10-OCT-2003 (rel. 42, Last annotation update)  
DE Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.  
RC TISSUE=Petal brain;  
RX MEDLINE=98324783; PubMed=9662406;  
RA Savukoski M., Kloockars T., Holmberg V., Santavuori P., Lander E.S.,  
RA Peltonen L.;  
RT "CLN5, a novel gene encoding a putative transmembrane protein mutated  
RT in Finnish variant late infantile neuronal ceroid lipofuscinosis.";  
RT Nat. Genet. 19:286-288 (1998).  
RN [2]  
SUBCELLULAR LOCATION, AND GLYCOSYLATION.  
RX MEDLINE=21968572; PubMed=11971870;  
RA Isoomppi J., Vesa J., Jalanko A., Peltonen L.;  
RT "Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5  
RT protein.";  
RL Hum. Mol. Genet. 11:885-891 (2002).  
CC -!- FUNCTION: Not known.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- PTM: Glycosylated.  
CC -!- DISEASE: Defects in CLN5 are the cause of Finnish variant late-  
CC infantile neuronal ceroid lipofuscinosis (VLINCL) [MIM:256731];  
CC also known as ceroid lipofuscinosis neuronal 5 (CLN5). VLINCL is a  
CC fatal childhood neurodegenerative disease characterized by  
CC progressive visual and mental decline, motor disturbance, epilepsy  
CC and behavioral changes. The first symptom is motor clumsiness,  
CC followed by progressive visual failure, mental and motor  
CC deterioration and later by myoclonia and seizures.  
CC -!- DATABASE: NAME=NCL CLN5;  
CC NOTE=Neural Ceroid Lipofuscinoses mutation db;  
CC WWW="http://www.ucl.ac.uk/ncl/CLN5.html".  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC EMBL; AF068227; AAC27614.1; -;  
CC DR GenBank; HGNC:2076; CLN5.  
CC DR MIM; 608102; -;

```

JR MIM; 256731; -
JR GO; GO:0016021; C: integral to membrane; TAS.
JR GO; GO:0008151; P: cell growth and/or maintenance; TAS.
JR Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
CM Disease mutation; Polymorphism; Epilepsy.
JR TRANSMEM 75
JR CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL)
JR VARIANT 279 279 D -> N (in VLIINCL).
JR VARIANT 368 368 /FTID=VAR_005137.
JR VARIANT K -> R (in dBSNP:1800209).
JR SEQUENCE 407 AA; 46339 MM; 449702DIDC9BFEE4 CRC64;
JR /FTID=VAR_005138.
JR /FTID=VAR_005138.

Query Match 100.0%; Score 154; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 3.5e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRGAGAGRAGRAGCAGATATLMLAVPQ 28
2 |||||
3 62 MRGAGAGRAGRAGCAGATATLMLAVPQ 89

RESULT 2
JR CBH_RALSO STANDARD; PRT; 126 AA.
JR Q8XZR2;
JR 28-FEB-2003 (Rel. 41, Created)
JR 28-FEB-2003 (Rel. 41, Last sequence update)
JR 28-FEB-2003 (Rel. 41, Last annotation update)
JR Protein crxb homolog.
JR CRXB OR RSC133 OR RS02855.
JR Ralstonia solanacearum (Pseudomonas solanacearum).
JR Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
JR Burkholderiaceae; Ralstonia.
JR NCBI_TaxID=305;
JR [1]
JR SEQUENCE FROM N.A.
JR STRAIN=GM11000;
JR MEDLINE=21681879; PubMed=11823852;
JR Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
JR Ariat M., Billault A., Broctier P., Camus J.C., Cattolico L.,
JR Chandler M., Choisme N., Clandel-Renard C., Cunnac S., Demange N.,
JR Gaupin C., Lavie M., Moisan A., Robert C., Saurin W., Schiek T.,
JR Siguiet P., Thebaud P., Whalen M., Wincker P., Levy M.,
JR Weissenbach J., Boucher C.A.;
JR "Genome sequence of the plant pathogen Ralstonia solanacearum";
JR Nature 415:497-502(2002).
JR -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
JR -1- SIMILARITY: Belongs to the crxb family.
JR -----
JR This SWISS-PROT entry is copyright. It is produced through a collaboration
JR between the Swiss Institute of Bioinformatics and the EMBL outstation -
JR the European Bioinformatics Institute. There are no restrictions on its
JR use by non-profit institutions as long as its content is in no way
JR modified and this statement is not removed. Usage by and for commercial
JR entities requires a license agreement (see http://www.isb-sib.ch/announce/
JR or send an email to license@isb-sib.ch).
JR -----
JR EMBL; AF646064; CAD15035.1; -
JR HAMAP; ME0454; -; 1.
JR InterPro; IPR003691; Campor_Crxb.
JR Pfam; PF02537; CRXB; 1.
JR TIGRFAVS; TIGR00494; crxb; 1.
JR TRANSMEM 5 24 POTENTIAL.
JR TRANSMEM 34 56 POTENTIAL.

```

```

JR TRANSMEM 68 90 POTENTIAL.
JR TRANSMEM 100 122 POTENTIAL.
JR SO SEQUENCE 126 AA; 13101 MM; C2443FBAB5C81CB3 CRC64;

Query Match 36.7%; Score 56.5; DB 1; Length 126;
Best Local Similarity 52.0%; Pred. No. 0.85;
Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;

4 GAGAGRAGRAGCAGATATLMLAVPQ 27
5 |||||
6 10 GAGAGRAGRAGCAGATATLMLAVPQ 32

RESULT 3
JR MTS_BOVIN STANDARD; PRT; 575 AA.
JR ID MTS_BOVIN
JR AC P03972;
JR DT 23-OCT-1986 (Rel. 02, Created)
JR DT 23-OCT-1986 (Rel. 02, Last sequence update)
JR DT 16-OCT-2001 (Rel. 40, Last annotation update)
JR DE Muellerian inhibiting factor precursor (MTS) (Anti-muellerian hormone)
JR DE (AMH) (Muellerian inhibiting substance).
JR GN AMH.
JR OS Bos taurus (Bovine).
JR OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JR OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
JR OC Bovidae; Bovinae; Bos.
JR NCBI_TaxID=9913;
JR [1]
JR SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
JR MEDLINE=86218082; PubMed=3754790;
JR Cate R.L., Mattaliano R.J., Hession C., Tizard R., Farber N.M.,
JR Cheung A., Nifra E.G., Frey A.Z., Gash D.J., Chow E.P., Fisher R.A.,
JR Bertonis J.W., Torres G., Wallner B.P., Ramachandran K.L.,
JR Regin R.C., Mangano T.F., McLaughlin D.T., Donahoe P.K.;
JR "Isolation of the bovine and human genes for Mullerian inhibiting
JR substance and expression of the human gene in animal cells.";
JR Cell 45:685-698(1986).
JR -1- FUNCTION: THIS GLYCOPROTEIN, PRODUCED BY THE SEROTOLI CELLS OF THE
JR TESTIS, CAUSES REGRESSION OF THE MUELLERIAN DUCT. IT ALSO IS ABLE
JR TO INHIBIT THE GROWTH OF TUMORS DERIVED FROM TISSUES OF MUELLERIAN
JR DUCT ORIGIN.
JR -1- SUBUNIT: Homodimer; disulfide-linked.
JR -1- MISCELLANEOUS: Although it does not compete with EGF for receptor
JR binding sites, MTS can inhibit the autophosphorylation of the EGF
JR receptor in vitro.
JR -1- SIMILARITY: Belongs to the TGF-beta family.
JR -----
JR This SWISS-PROT entry is copyright. It is produced through a collaboration
JR between the Swiss Institute of Bioinformatics and the EMBL outstation -
JR the European Bioinformatics Institute. There are no restrictions on its
JR use by non-profit institutions as long as its content is in no way
JR modified and this statement is not removed. Usage by and for commercial
JR entities requires a license agreement (see http://www.isb-sib.ch/announce/
JR or send an email to license@isb-sib.ch).
JR -----
JR EMBL; M13151; AAA98765.1; -
JR PIR; A01398; WEFOM.
JR InterPro; IPR006799; AMH_N.
JR InterPro; IPR002400; GF_CysKnot.
JR InterPro; IPR001839; TGFb.
JR Pfam; PF04709; AMH_N; 1.
JR Pfam; PF00019; TGF-beta; 1.
JR PRINTS; PR00439; GRCYSKNOT.
JR PRODOM; PD000357; TGFb; 1.
JR SMART; SM00204; TGFb; 1.
JR PROSITE; PS00250; TGF_BETA_1; 1.
JR Growth factor; Glycoprotein; Gonadal differentiation; Signal.
JR STGNAL 1 17
JR PROPEP 16 24
JR FT 25 575 POTENTIAL.
JR FT 477 541 MUELLERIAN INHIBITING FACTOR.
JR FT 503 572 BY SIMILARITY.

```

```

T DISUFIID 507 574 BY SIMILARITY.
T DISUFIID 540 540 INTERCHAIN (BY SIMILARITY).
T CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
Q SEQUENCE 575 AA; 60623 MW; 8928B9C1ACB5A8 CRC64;

Query Match 34.1%; Score 52.5; DB 1; Length 575;
Best Local Similarity 44.0%; Pred. No. 11;
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

Y 4 GAGAGRGASGCMWALMLTAVVPG 28
b 41 GSGALIFQCAWMPPLSLWLV---PG 62

RESULT 4
ZDI_RAT STANDARD; PRT; 641 AA.
ZDI_RAT
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Frizzled 1 precursor (Frizzled-1) (Fz-1) (Fz1).
FZD1.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
STRAIN=Sprague-Dawley; TISSUE=Osteosarcoma;
MEDLINE=93094228; PubMed=1334084;
Chan S.D.H., Karpf D.B., Fowlkes M.E., Hooks M., Bradley M.S.,
Vuong V., Bamhidi T., Liu M.Y.C., Arnard C.D., Stremler G.J.,
Nissenen R.A.;
"Two homologs of the Drosophila polarity gene frizzled (Fz) are widely
expressed in mammalian tissues.";
J. Biol. Chem. 267:25202-25207 (1992).
[2]
COUPLING TO BETA-CATENIN PATHWAY.
MEDLINE=99324245; PubMed=10395542;
Sheldahl L.C., Park M., Malbon C.C., Moon R.T.;
"Protein kinase C is differentially stimulated by Wnt and Frizzled
homologs in a G-protein-dependent manner.";
Curr. Biol. 9:695-698 (1999).
-1- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors
are coupled to the beta-catenin canonical signaling pathway, which
leads to the activation of dishevelled proteins, inhibition of
GSK-3 kinase, nuclear accumulation of beta-catenin and activation
of Wnt target genes. A second signaling pathway involving PKC and
calcium fluxes has been seen for some family members, but it is
not yet clear if it represents a distinct pathway or if it can be
integrated in the canonical pathway, as PKC seems to be required
for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem
to involve interactions with G-proteins. May be involved in
transduction and intercellular transmission of polarity
information during tissue morphogenesis and/or in differentiated
tissues. Activation by Wnt induces expression of beta-catenin
target genes.
-1- SUBCELLULAR LOCATION: Integral membrane protein.
-1- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney,
liver, uterus, ovary and heart. Lower levels seen in brain and
intestine. Extremely low in calvaria, mammary glands and testis.
-1- DEVELOPMENTAL STAGE: Expressed predominantly in neonatal tissues,
at lower levels in adult.
-1- DOMAIN: Lys-Thr-X-X-X-Tip motif is involved in the activation of
the Wnt/beta-catenin signaling pathway (By similarity).
-1- DOMAIN: The Fz domain is involved in binding with Wnt ligands (By
similarity).
-1- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor
family.
-1- SIMILARITY: Contains 1 frizzled (FZ) domain.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02529; AAA41173.1; .
DR InterPro; IPR000539; Frizzled.
DR InterPro; IPR000024; Fz domain.
DR InterPro; IPR000832; GPCR secretin.
DR Pfam; PF01534; Frizzled; 1.
DR Pfam; PF01392; Fz; 1.
DR PRINTS; PRO0489; FRIZZLED.
DR SMART; SM00063; FRI; 1.
DR PROSITE; PS0038; FZ; 1.
DR PROSITE; PS0261; G-PROTEIN RECEPTOR; Transmembrane;
DR Multigene family; G-protein coupled receptor; Transmembrane;
DR Developmental protein; Wnt signaling pathway; Glycoprotein; Signal.
DR SIGNAL.
FT CHAIN 1 68
FT DOMAIN 69 641
FT TRANSMEM 317 337
FT DOMAIN 338 348
FT TRANSMEM 349 369
FT DOMAIN 370 396
FT TRANSMEM 397 417
FT DOMAIN 418 439
FT TRANSMEM 440 460
FT DOMAIN 461 483
FT TRANSMEM 484 504
FT TRANSMEM 505 530
FT TRANSMEM 531 551
FT DOMAIN 552 595
FT TRANSMEM 596 616
FT DOMAIN 617 641
FT DOMAIN 106 224
FT SITE 85 90
FT SITE 619 624
FT SITE 639 641
FT CARBOHYD 125 125
FT CARBOHYD 225 225
SQ SEQUENCE 641 AA; 71027 MW; D8E2C11B81B8B6 CRC64;

Query Match 33.8%; Score 52; DB 1; Length 641;
Best Local Similarity 55.6%; Pred. No. 14;
Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

QY 5 AGAARGRA-SWCM---LALMLAVV 27
Db 38 AGHRRPRAHRCMARGLLMLLEAP 64

RESULT 5
PREL_YEAST STANDARD; PRT; 686 AA.
AC P32791.
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ferric reductase transmembrane component 1 precursor (EC 1.16.1.7)
DE (Ferric-chelate reductase 1).
GN PREL OR YIR214W OR D8167.2.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
ON NCBI_TaxID=4932;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F113;
RX MEDLINE=92237270; PubMed=1570306;
RA Dancis A., Roman D.G., Anderson G.J., Himebusch A.G., Klausner R.D.;

```







```
DR EMBL, AF123757, AAF13115.1, -.
DR EMBL, AF123758, AAF13116.1, -.
DR EMBL, AF123759, AAF13117.1, -.
DR EMBL, AF123760, AAF13118.1, -.
DR EMBL, AF123761, AAF13119.1, -.
DR EMBL, BC007725, AA077725.1, -.
DR GenBank, HGNC:2079, CLN8.
DR MIM, 607837, -.
DR GO, GO:0005783, C:Endoplasmic reticulum; TAS.
DR GO, GO:0005793, C:ER-Golgi intermediate compartment; TAS.
DR GO, GO:0016021, C:Integral to membrane; TAS.
DR GO, GO:0007399, P:neurogenesis; TAS.
DR InterPro, IPR006634, TLC.
DR SMART, SM00724, TLC; 1.
DR PROSITE, PS50922, TLC; 1.
KW Transmembrane; Endoplasmic reticulum; Neuronal ceroid lipofuscinosis;
KW Disease mutation; Polymorphism; Epilepsy.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 62 84 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT DOMAIN 62 262 TLC.
FT SITE 283 286 ER-RETRIEVAL SIGNAL.
FT VARIANT 24 24 R->G (in EPMR).
FT VARIANT 155 155 /FTID=VAR_013174.
FT MUTAGEN 283 284 KK->RR: LOCALIZES TO THE GOLGI COMPLEX.
FT CONFLICT 225 225 N->S (IN REF. 2).
SQ SEQUENCE 286 AA; 32842 MW; 0BAAEBCA516D1DC CRC64;

Query Match 31.8%; Score 49; DB 1; Length 286;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 AGAARGASMCW 16
DB 93 ADKARGQGNWCM 104

RESULT 9
PEXD CAEEL STANDARD; PRT; 330 AA.
AC 019951;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable peroxisomal membrane protein PEX13 (Peroxin-13).
GN P32A5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Pauley A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Component of the peroxisomal translocation machinery
CC with PEX14 and PEX17. Functions as a docking factor for the
CC predominantly cytoplasmic PPS1 receptor (PPS10/PEX5) (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (By
CC similarity).
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U20864, AAC46665.2; -.
DR HSSP, P00519, IAB1.
DR WormPep, F32A5.6; CE28290.
DR InterPro, IPR007223, Peroxin-13_N.
DR InterPro, IPR001452, SH3.
DR Pfam, PF04088, Peroxin-13_N; 1.
DR Pfam, PF00018, SH3; 1.
DR PRINTS, PR00452, SH3DOMAIN.
DR PRODOM, PD000066, SH3; 1.
DR SMART, SM00326, SH3; 1.
DR PROSITE, PS50002, SH3; 1.
KW Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.
FT DOMAIN 1 192 LUMENAL (POTENTIAL).
FT TRANSMEM 193 213 POTENTIAL.
FT DOMAIN 214 330 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 235 300 SH3.
SQ SEQUENCE 330 AA; 35765 MW; 7DD647F07A0403 CRC64;

Query Match 31.8%; Score 49; DB 1; Length 330;
Best Local Similarity 43.5%; Pred. No. 19;
Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

OY 6 GAARGASMCWALMLAVVG 28
DB 187 GATRTPASVWPALFWVAIG 209

RESULT 10
LNT_SYNY3 STANDARD; PRT; 519 AA.
AC P74055;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein N-acyltransferase (EC 2.3.1.-) (AIP N-acyltransferase).
GN LNT OR SLR0819.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairu K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yanada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: Transfers the fatty acyl group on membrane lipoproteins
CC (by similarity).
CC -1- PATHWAY: Lipoproteins biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the apolipoprotein N-acyltransferase
CC family.
CC -1- SIMILARITY: Contains 1 CN hydrolase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```

1000 -1- TISSUE SPECIFICITY: Ubiquitously expressed with high expression in
1001 testis, moderate expression in the pancreas, spleen, prostate,
1002 small intestine and colon. Very low expression is seen in brain
1003 and skeletal muscle.
1004 -1- SIMILARITY: Contains 3 MIR domains.
1005 -----
1006 This SWISS-PROT entry is copyright. It is produced through a collaboration
1007 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
1008 the European Bioinformatics Institute. There are no restrictions on its
1009 use by non-profit institutions as long as its content is in no way
1010 modified and this statement is not removed. Usage by and for commercial
1011 entities requires a license agreement (See http://www.isb-sib.ch/announce/
1012 or send an email to license@isb-sib.ch).
1013 -----
1014 EMBL; AB043007; BAB18277.1; -
1015 EMBL; AF277316; AAK69113.1; -
1016 EMBL; BC006248; AAH06248.1; -
1017 PIR; JC7587; JC7587; SDF2L1.
1018 Genew; HGNC:10676; SDF2L1.
1019 MIR; 607551; -
1020 InterPro; IPR000886; ER_target_S.
1021 InterPro; IPR003608; MIR.
1022 Pfam; PF02815; MIR; 3.
1023 SMART; SMO0472; MIR; 3.
1024 PROSITE; PS00014; ER_TARGET; 1.
1025 PROSITE; PS00919; MIR; 3.
1026 Endoplasmic reticulum; Signal; Repeat.
1027 SIGNAL; 1; 28; POTENTIAL.
1028 CHAIN; 29; 221; STROMAL CELL-DERIVED FACTOR 2-LIKE
1029 PROTEIN 1.
1030 DOMAIN; 33; 87; MIR 1.
1031 DOMAIN; 95; 150; MIR 2.
1032 SITE; 151; 205; MIR 3.
1033 SITE; 218; 221; PREVENT SECRETION FROM ER (POTENTIAL).
1034 CONFLICT; 94; 94; C -> R (IN REF. 2 AND 3).
1035 CONFLICT; 162; 162; L -> F (IN REF. 2 AND 3).
1036 SEQUENCE; 221 AA; 23511 MW; 23EBBE14BEE9E2BF CRC64;
1037
1038 Query Match 31.2%; Score 48; DB 1; Length 221;
1039 Best Local Similarity 50.0%; Pred. No. 18;
1040 Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
1041
1042 Y 7 AARGASWCWALALMLAVVG 28
1043 | | | | | | | | | |
1044 | | | | | | | | | |
1045 ID P1C1_AGR5 STANDARD; PRT; 233 AA.
1046 FC P29112;
1047 FC 01-DEC-1992 (Rel. 24, Created)
1048 FC 01-DEC-1992 (Rel. 24, Last sequence update)
1049 FC 28-FEB-2003 (Rel. 41, Last annotation update)
1050 FC 24.9 kDa protein in p1CA locus (ORF1).
1051 FC ATU3128 OR AGR_L 3363.
1052 FC Agrobacterium tumefaciens (strain C58 / ATCC 33970), and
1053 Agrobacterium tumefaciens.
1054 FC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
1055 FC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1056 FC NCBI_TaxID=176299; 358;
1057 FC [1]
1058 FC SEQUENCE FROM N.A.
1059 FC STRAIN=C58 / ATCC 33970;
1060 FC MEDLINE=21608550; Pubmed=11743193;
1061 FC Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
1062 FC Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
1063 FC Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D., Sr.,
1064 FC Chapman P., Clendenning J., Decherage G., Gillet M., Grant C.,
1065 FC Kutyavin T., Levy R., Li M.-d., McClelland E., Palmieri A.,
1066 FC Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
1067 FC Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
1068 FC Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

```

```

1069 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
1070 RA Neeter E.W.;
1071 RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
1072 RA C58.";
1073 RA Science 294:2317-2323(2001).
1074 RL [2]
1075 RN SEQUENCE FROM N.A.
1076 RC STRAIN=C58 / ATCC 33970;
1077 RC MEDLINE=21608551; Pubmed=11743194;
1078 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
1079 RA Houmli B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L.,
1080 RA Houmli K., Gordon J., Vaudin M., Iartchouk O., Bpp A., Liu F.,
1081 RA Wolan C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
1082 RA Flanagan C., Crowell C., Garson J., Lomo C., Sear C., Strub G.,
1083 RA Cielo C., Slater S.,
1084 RT "Genome sequence of the plant pathogen and biotechnology agent
1085 RT Agrobacterium tumefaciens C58.";
1086 RL Science 294:2323-2328(2001).
1087 RN [3]
1088 RP SEQUENCE FROM N.A.
1089 RC STRAIN=A136;
1090 RC MEDLINE=91317730; Pubmed=1860822;
1091 RA Rong L., Karcher S.-J., Gelvin S.B.;
1092 RT "Genetic and molecular analyses of p1CA, a plant-inducible locus on
1093 RT the Agrobacterium tumefaciens chromosome.";
1094 CC J. Bacteriol. 173:5110-5120(1991).
1095 -1- FUNCTION: SEEM TO REGULATE THE SURFACE PROPERTIES OF THE
1096 BACTERIUM IN THE PRESENCE OF PLANT CELLS OR PLANT CELL EXTRACTS.
1097 MUTATIONS IN THIS PROTEIN ARE RESPONSIBLE FOR AN INCREASED
1098 AGGREGATION OF THE BACTERIA IN THE PRESENCE OF PEA ROOT CAP
1099 CELLS.
1100 CC -1- INDUCTION: By certain acidic polysaccharides found in carrot root
1101 CC extract. This induction may be regulated by the polygalacturonase.
1102 CC
1103 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
1104 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
1105 CC the European Bioinformatics Institute. There are no restrictions on its
1106 CC use by non-profit institutions as long as its content is in no way
1107 CC modified and this statement is not removed. Usage by and for commercial
1108 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
1109 CC or send an email to license@isb-sib.ch).
1110 CC -----
1111 CC EMBL; AE009243; AAL43944.1; ALT_INT.
1112 CC DR EMBL; AE008370; AAK90258.1; ALT_INT.
1113 CC DR EMBL; M62814; AAA22103.1; -
1114 CC DR PIR; B40364; B40364.
1115 CC Complete proteome.
1116 SQ SEQUENCE 233 AA; 24935 MW; 4ED00E8C180195F7 CRC64;
1117
1118 Query Match 31.2%; Score 48; DB 1; Length 233;
1119 Best Local Similarity 41.7%; Pred. No. 19;
1120 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
1121
1122 QY 2 RRGAGARGASWCWALALMLAV 25
1123 | | | | | | | | | |
1124 | | | | | | | | | |
1125 DB 50 QEWANPASKSPAIWARAVGLAM 73
1126
1127 RESULT 14
1128 NIDO MOUSE
1129 ID NIDO_MOUSE STANDARD; PRT; 1245 AA.
1130 AC P10493;
1131 DT 01-APR-1990 (Rel. 14, Created)
1132 DT 01-APR-1990 (Rel. 14, Last sequence update)
1133 DT 15-MAR-2004 (Rel. 43, Last annotation update)
1134 DE Nidogen precursor (Enactin).
1135 GN NID OR NIDI OR ENT.
1136 OS Mus musculus (Mouse).
1137 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
1138 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1139 NCBI_TaxID=10090;
1140 RN [1]
1141 SEQUENCE FROM N.A., AND SEQUENCE OF 29-40.

```

MEDLINE=89079780; PubMed=3264556;  
Durkin M.E.; Chakravarti S.; Bartos B.B.; Liu S.H.; Friedman R.L.;  
Chung A.E.;  
"Amino acid sequence and domain structure of entactin. Homology with  
epidermal growth factor precursor and low density lipoprotein  
receptor";  
J. Cell Biol. 107:2749-2756(1988).  
[2]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
MEDLINE=89231638; PubMed=2496973;  
Mann K., Deutzmann R., Ammalleit W., Timpl R., Raimondi L.,  
Yamada Y., Pan T.-C., Conway D., Chu M.-L.;  
"Amino acid sequence of mouse nidogen, a multidomain basement  
membrane protein with binding activity for laminin, collagen IV and  
cells";  
EMBO J. 8:65-72(1989).  
[3]  
SEQUENCE OF 1-251 FROM N.A.  
STRAIN=BALB/c; TISSUE=Liver;  
MEDLINE=94040771; PubMed=8224873;  
Durkin M.E., Liu S.H., Reing J., Chung A.E.;  
"Characterization of the 5' end of the mouse Ent gene encoding the  
basement membrane protein, entactin.";  
Gene 132:261-266(1993).  
[4]  
SEQUENCE OF 1207-1245 FROM N.A.  
STRAIN=C57BL/6J X CBA/J;  
MEDLINE=95324932; PubMed=7601446;  
Durkin M.E., Wewer U.M., Chung A.E.;  
"Exon organization of the mouse entactin gene corresponds to the  
structural domains of the polypeptide and has regional homology to  
the low-density lipoprotein receptor gene.";  
Genomics 26:219-228(1995).  
[5]  
PARTIAL SEQUENCE.  
MEDLINE=86192477; PubMed=3084254;  
Paulsson M., Deutzmann R., Dziadek M., Nowack H., Timpl R., Weber S.;  
Engel J.;  
"Purification and structural characterization of intact and  
fragmented nidogen obtained from a tumor basement membrane.";  
Eur. J. Biochem. 156:467-478(1986).  
[6]  
CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.  
MEDLINE=93316903; PubMed=8326911;  
Fujisawa S., Shinkai H., Mann K.; Timpl R.;  
"Structure and localization of O- and N-linked oligosaccharide chains  
on basement membrane protein nidogen.";  
Matrix 13:215-222(1993).  
[7]  
INTERACTION WITH FBNI.  
MEDLINE=97446166; PubMed=9299350;  
Adam S., Goehring W., Wiedemann H., Chu M.-L., Timpl R., Koska G.;  
"Binding of fibulin-1 to nidogen depends on its C-terminal globular  
domain and a specific array of calcium-binding epidermal growth  
factor-like (EGF) modules.";  
J. Mol. Biol. 272:226-236(1997).  
[8]  
INTERACTION WITH FBNI.  
MEDLINE=21474010; PubMed=11589703;  
Ries A., Goehring W., Fox U.W., Timpl R., Sasaki T.;  
"Recombinant domains of mouse nidogen-I and their binding to basement  
membrane proteins and monoclonal antibodies.";  
Eur. J. Biochem. 268:5119-5128(2001).  
-1- FUNCTION: Sulfated glycoprotein which is widely distributed in  
basement membranes and that is tightly associated with laminin.  
Also binds to collagen IV. It probably has a role in cell-  
extracellular matrix interactions.  
-1- SUBUNIT: Interacts with FBNI.  
-1- SUBCELLULAR LOCATION: Basement membranes.  
-1- PTM: N- and O-glycosylated.  
-1- SIMILARITY: Contains 6 EGF-like domains.  
-1- SIMILARITY: Contains 1 fibroglobulin type-I domain.  
-1- SIMILARITY: Contains 5 IGD-receptor WWTD domains.

	CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC
	CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed.
	CC	Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licenses@isb-sib.ch).
	CC	
	DR	EMBL; X14194; CAA32408.1; -. DR
	DR	EMBL; X14480; CAA32642.1; -. DR
	DR	EMBL; L17324; AAA77652.1; JOINED. DR
	DR	EMBL; L17323; AAA77652.1; JOINED. DR
	DR	EMBL; X83093; CAA58148.1; -. PIR; S02730; MMSMND. DR
	DR	PDB; 1GL4; 28-NOV-01. PDB; 1H4U; 28-FEB-03. DR
	DR	MGI; MG1; 97342; NidI. DR
	DR	GO; GO:0005604; C:basesment membrane; IDA. DR
	DR	GO; GO:0005515; F:protein binding; IPT. DR
	DR	GO; GO:0007160; P:cell-matrix adhesion; IDA. DR
	DR	InterPro; IPR000152; Asx_hydroxy1_S. DR
	DR	InterPro; IPR001881; EGF_Ca. DR
	DR	InterPro; IPR006209; EGF_like. DR
	DR	InterPro; IPR006605; GZF. DR
	DR	InterPro; IPR003017; GFP_like. DR
	DR	InterPro; IPR000033; Idl_receptor_rep. DR
	DR	InterPro; IPR003886; Nidogen_ext. DR
	DR	InterPro; IPR000716; Thyroglobulin_1. DR
	DR	Pfam; PF00008; EGF; 6. DR
	DR	Pfam; PF00058; Idl_recept_b; 3. DR
	DR	Pfam; PF00086; thyroglobulin_1; 1. DR
	DR	SMART; SMART0179; EGF_CA; 2. DR
	DR	SMART; SMART682; GZF; 1. DR
	DR	SMART; SMART0135; LY; 5. DR
	DR	SMART; SMART0539; NIDO; 1. DR
	DR	SMART; SMART0211; TY; 1. DR
	DR	PROSITE; PS00010; ASX_HYDROXYL; 3. DR
	DR	PROSITE; PS00022; EGF_1; 1. DR
	DR	PROSITE; PS01186; EGF_2; 4. DR
	DR	PROSITE; PS50026; EGF_3; 5. DR
	DR	PROSITE; PS01187; EGF_CA; 2. DR
	KW	Basement membrane; Extracellular matrix; Glycoprotein; Sulfation; Signal; Calcium-binding; Repeat; Egf-like domain; Cell adhesion; 3d-structure. KM
	FT	SIGNAL 1..28 FT
	FT	CHAIN 29..1245 FT
	FT	DOMAIN 29..667 II (LARGER GLOBULAR DOMAIN). FT
	FT	DOMAIN 668..919 III (SMALLER GLOBULAR DOMAIN). FT
	FT	DOMAIN 920..1245 EGF-LIKE 1. FT
	FT	DOMAIN 384..424 EGF-LIKE 2. FT
	FT	DOMAIN 666..707 EGF-LIKE 3. FT
	FT	DOMAIN 708..749 EGF-LIKE 4. FT
	FT	DOMAIN 756..799 EGF-LIKE 5. FT
	FT	DOMAIN 800..838 EGF-LIKE 6. FT
	FT	DOMAIN 870..917 THYROGLOBULIN TYPE-I. FT
	FT	DOMAIN 987..1028 LDL-RECEPTOR YWTD MOTIF 1. FT
	FT	DOMAIN 1030..1071 LDL-RECEPTOR YWTD MOTIF 2. FT
	FT	DOMAIN 1073..1116 LDL-RECEPTOR YWTD MOTIF 3. FT
	FT	DOMAIN 1122..1161 LDL-RECEPTOR YWTD MOTIF 4. FT
	FT	DOMAIN 1206..1242 SULFATION (POTENTIAL). FT
	FT	MOD_RES 290..290 SUFATON (POTENTIAL). FT
	FT	MOD_RES 295..295 BY SIMILARITY. FT
	FT	DISULEID 670..683 BY SIMILARITY. FT
	FT	DISULEID 677..693 BY SIMILARITY. FT
	FT	DISULEID 695..706 BY SIMILARITY. FT
	FT	DISULEID 712..725 BY SIMILARITY. FT
	FT	DISULEID 719..734 BY SIMILARITY. FT
	FT	DISULEID 736..748 BY SIMILARITY. FT
	FT	TIT 760..775 BY SIMILARITY. FT
	FT	DISULEID 767..785 BY SIMILARITY. FT

```

T DISULFID 787 798 BY SIMILARITY.
T DISULFID 804 815 BY SIMILARITY.
T DISULFID 809 824 BY SIMILARITY.
T DISULFID 826 837 BY SIMILARITY.
T DISULFID 1210 1221 BY SIMILARITY.
T DISULFID 1217 1230 BY SIMILARITY.
T DISULFID 1232 1241 BY SIMILARITY.
T SITE 700 702 CELL ATTACHMENT SITE.
T CARBOHYD 187 187 N-LINKED (GLCNAC. . .).
T CARBOHYD 299 299 O-LINKED.
T CARBOHYD 331 331 O-LINKED.
T CARBOHYD 337 337 O-LINKED.
T CARBOHYD 345 345 O-LINKED.
T CARBOHYD 348 348 O-LINKED (PARTIAL. . .).
T CARBOHYD 415 415 N-LINKED (GLCNAC. . .).
T CARBOHYD 920 920 O-LINKED.
T CARBOHYD 933 933 O-LINKED.
T CONFLICT 170 170 P -> L (IN REF. 2).
T CONFLICT 659 659 R -> K (IN REF. 2).
T CONFLICT 967 967 R -> A (IN REF. 2).
Q SEQUENCE 1245 AA; 136622 MW; 8FBE276F29BE6D2 CRC64;

```

```

Query Match 31.2%; Score 48; DB 1; Length 1245;
Best Local Similarity 58.8%; Pred. No. 82;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Y 12 ASWCWALALMLAVPG 28
b 10 AMWTMLLQLLVGPG 26

```

```

ESULT 15
JGB_ECOLI

```

```

D YJGB_ECOLI STANDARD; PRT; 339 AA.
P27250; P76812;

```

```

T 01-AUG-1992 (Rel. 23, Created)
T 01-FEB-1995 (Rel. 31, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E Hypochemical zinc-type alcohol dehydrogenase-like protein yjgb.
N YJGB OR B4269.
S Escherichia coli.
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
C Enterobacteriaceae; Escherichia.
C NCBI_TaxID=562;

```

```

X 1] SEQUENCE FROM N.A.

```

```

C STRAIN=B.

```

```

A Pucci M.J., Discotto L.F., Dougherty T.J.;
Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.

```

```

L [2]

```

```

P SEQUENCE FROM N.A.

```

```

C STRAIN=K12 / MG1655;

```

```

X MEDLINE=95334362; PubMed=7610040;
Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,

```

```

A Blatner F.R.;
"Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes."

```

```

L Nucleic Acids Res. 23:2105-2119(1995).

```

```

C -1- COPACTOR: Binds 2 zinc ions per subunit (By similarity).

```

```

C -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
family.

```

```

C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).

```

```

R EXBL: M96355; AAA72122.1;
R EMBL: U14003; AAA97166.1; ALT_INIT.
R EMBL: AE000497; AAC77226.1; ALT_INIT.

```

```

DR EcGene; EG11436; Yjgb.
DR InterPro; IPR002128; Adh_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; Adh_zinc_N; 1.
DR PROSITE; PS00059; Adh_ZINC; 1.
KW Hypochemical protein; Oxidoreductase; Zinc; Metal-binding;
KW Complete proteome.
FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 63 63 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 96 96 ZINC 2 (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 102 102 ZINC 2 (BY SIMILARITY).
FT METAL 110 110 ZINC 2 (BY SIMILARITY).
FT METAL 152 152 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 331 339 YRVLRADF -> TAWC (IN REF. 1).
SQ SEQUENCE 339 AA; 36502 MW; 0854DDEFA16B9EE CRC64;

```

```

Query Match 30.5%; Score 47; DB 1; Length 339;
Best Local Similarity 45.0%; Pred. No. 36;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 MNRGAGARGRASWCWALAI 20
Db 119 MNRGCFARLRADWQVPIPL 138

```

```

Search completed: June 8, 2004, 14:04:38
Job time : 3.16667 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 13:59:08 ; Search time 5.17222 Seconds

(without alignments)  
1708.068 Million cell updates/sec

Title: US-10-010-050A-2\_COPY\_1\_28  
Effect score: 154  
Sequence: 1 MRRGGAARGRASMCALALMLAVPG 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Search: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Set-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPRENBL\_25.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rickettsia:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66	42.9	606	17 Q9HMT4	Q9HMT4 halobacter.
2	59	38.3	412	16 Q92VB2	Q92VB2 rhizobium n
3	59	38.3	588	16 Q82G94	Q82G94 streptomyce
4	58	37.7	146	11 Q8CEK6	Q8CEK6 mus musculu
5	57.5	37.3	343	16 Q8G419	Q8G419 bifidobacte
6	57	37.0	366	10 Q84X72	Q84X72 chlamydomon
7	56	36.4	214	16 Q69464	Q69464 mycobacter.
8	56	36.4	485	16 Q88SV3	Q88SV3 lactobacill.
9	55.5	36.0	412	10 Q7XSY0	Q7XSY0 oryza sativ
10	55.5	36.0	695	16 Q8XW44	Q8XW44 ralscorta s
11	55	35.7	147	10 Q7X917	Q7X917 oryza sativ
12	55	35.7	304	16 Q7WFK8	Q7WFK8 bordetella
13	55	35.7	304	16 Q7W451	Q7W451 bordetella
14	55	35.7	707	10 Q92R46	Q92R46 chlamydomon
15	55	35.7	3019	12 Q92529	Q92529 hepatitis c
16	54.5	35.4	205	16 Q8PDR8	Q8PDR8 xanthomonas

17	54.5	35.4	435	10 Q9XE69	Q9XE69 sorghum bic
18	54.5	35.4	836	16 Q89HK3	Q89HK3 bradyrhizob
19	54	35.1	214	10 Q9XEO6	Q9XEO6 sorghum bic
20	54	35.1	232	16 Q9PEH4	Q9PEH4 xyella fas
21	54	35.1	232	16 Q87EH4	Q87EH4 xyella fas
22	54	35.1	266	2 Q8L311	Q8L311 vitreoscill
23	53.5	34.7	126	2 Q8W997	Q8W997 alcaligenes
24	53	34.4	83	9 Q7Y4J5	Q7Y4J5 streptococc
25	53	34.4	160	16 Q82HW2	Q82HW2 streptomyce
26	53	34.4	224	2 Q9EY24	Q9EY24 xanthomonas
27	53	34.4	428	16 Q8UD14	Q8UD14 agrobacteri
28	52.5	34.1	201	16 Q8PE05	Q8PE05 xanthomonas
29	52.5	34.1	333	16 Q8NRA3	Q8NRA3 corynebacte
30	52.5	34.1	334	2 Q9KML7	Q9KML7 corynebacte
31	52	33.8	119	10 Q7XTZ2	Q7XTZ2 oryza sativ
32	52	33.8	455	16 Q88D16	Q88D16 pseudomonas
33	52	33.8	658	4 Q9BX78	Q9BX78 homo sapien
34	52	33.8	667	4 Q9BX79	Q9BX79 homo sapien
35	52	33.8	667	4 Q8TB21	Q8TB21 homo sapien
36	52	33.8	708	4 Q7Z3U9	Q7Z3U9 homo sapien
37	51.5	33.4	543	10 Q94DP0	Q94DP0 oryza sativ
38	51.5	33.4	1146	13 Q80584	Q80584 gallus gall
39	51.5	33.4	1252	10 Q7XU28	Q7XU28 oryza sativ
40	51.5	33.4	1252	10 Q8GU77	Q8GU77 oryza sativ
41	51	33.1	102	16 Q7UB52	Q7UB52 shigella fl
42	51	33.1	248	16 Q8PF85	Q8PF85 xanthomonas
43	51	33.1	410	16 Q92U79	Q92U79 rhizobium m
44	51	33.1	456	16 Q93JK1	Q93JK1 streptomyce
45	51	33.1	463	10 Q8H053	Q8H053 oryza sativ

## ALIGNMENTS

RESULT 1	ID	PRELIMINARY:	PRT:	606 AA.
Q9HMT4	Q9HMT4			
AC	Q9HMT4			
AD	Q9HMT4			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	Sulfate transport system permease protein.			
GN	CYS11 OR YNG2396G.			
OS	Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).			
OC	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=64091;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20504483; PubMed=11016950;			
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,			
RA	Shukla H.D., Laskey S.R., Balla N.S., Thorsson V., Sirogna T.,			
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,			
RA	Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,			
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,			
RA	Isehaberger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,			
RA	Alan M., Freitas T., Hou S., Daniels C.T., Dennis P.P., Omer A.D.,			
RA	Bhardt H., Lowe T.W., Liang P., Riley M., Hood L., Dassarma S.;			
RT	"Genome sequence of Halobacterium species NRC-1".			
RT	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).			
CC	-1- FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT			
CC	SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE			
CC	SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	-1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-			
CC	PROTEIN-DEPENDENT TRANSPORT SYSTEMS.			
DR	EMBL; AE005121; AAC20487.1; -.			
DR	PIR; C84390; C84390.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0003700; F:transcription factor activity; IEA.			
DR	GO; GO:0005215; F:transporter activity; IEA.			
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			



DR InterPro; IPR000515; BPD\_transp.  
DR InterPro; IPR000847; HTH\_LysR.  
DR Pfam; PF00528; BPD\_transd. 2.  
DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER; 1.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
DR Transmembrane; Transp; Complete proteome.  
KW SEQUENCE 606 AA; 63066 MW; 2E64B9A20298A28C CRC64;

Query Match 42.9%; Score 66; DB 17; Length 606;  
Best Local Similarity 53.8%; Pred. No. 1.1;  
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 3 RGAGARASWCWALALMLAVPG 28  
DB 508 RALGASRGALMDVLPVMEGVAG 533

RESULT 2  
Q92VB2 PRELIMINARY; PRT; 412 AA.  
AC Q92VB2;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE Hypothetical membrane protein SMD21292.  
GN RB0796 OR SMD21292.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
NCBI\_TaxID=382;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396508; PubMed=11481431;  
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Gouzy J.,  
RA Golding B., Pecher A.;  
RT "The complete sequence of the 1,683-Rb pSymb megaplasmid from the N2-  
fixing endosymbiont Sinorhizobium meliloti";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).  
DR EMBL; AL603644; CAC49196.1; -.  
DR PIR; D95941; D95941.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR000345; CYC\_heme\_BS.  
DR PROSITE; PS00190; CYTOCHROME\_C7.1.  
KW Plasmid; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 412 AA; 46126 MW; 8088B8E0C66F8838 CRC64;

Query Match 38.3%; Score 59; DB 16; Length 412;  
Best Local Similarity 37.0%; Pred. No. 6.9;  
Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY 7 AARGRASWCW-----LALMLAVP 27  
DB 276 ANKGRPTWTATVLTALIMLSTVP 302

RESULT 3  
Q82G94 PRELIMINARY; PRT; 588 AA.  
AC Q82G94;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN SAV4004.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomycetes.  
NCBI\_TaxID=33903;

EN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinoue M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22608306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinoue M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:526-531 (2003).  
DR EMBL; AP005037; BAC71716.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 588 AA; 63428 MW; D8DC5F23BE49A321 CRC64;

Query Match 38.3%; Score 59; DB 16; Length 588;  
Best Local Similarity 29.5%; Pred. No. 9.5;  
Matches 13; Conservative 5; Mismatches 4; Indels 22; Gaps 1;

QY 3 RGAGAR-----GRASWCWALALMLA 24  
DB 167 RGASSRTKLYATKNTLIGVSGRYGDKSSWSMAGLITLA 210

RESULT 4  
Q8CEK6 PRELIMINARY; PRT; 146 AA.  
ID Q8CEK6;  
AC Q8CEK6;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN 9330132005RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium;  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK020369; BAC25627.1; -.  
DR MED; MG1:1924802; 9330132005RIK.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 146 AA; 16423 MW; AF4B10EB630CB6D5 CRC64;

Query Match 37.7%; Score 58; DB 11; Length 146;  
Best Local Similarity 52.4%; Pred. No. 3.7;  
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 RRGAGARASWCWALALIM 22  
DB 44 RGPAGRRSQPSSCWGLIM 64

RESULT 5  
Q8G419 PRELIMINARY; PRT; 343 AA.  
ID Q8G419

08G419:  
01-MAR-2003 (TREMBlrel. 23, Created)  
01-MAR-2003 (TREMBlrel. 23, last sequence update)  
01-MAR-2003 (TREMBlrel. 23, last annotation update)  
Hypothetical protein.  
B1576.  
Bifidobacterium longum.  
Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
Bifidobacteriaceae; Bifidobacterium.  
NCBI\_TaxID=216816;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=NCC 2705;  
MEDLINE=22294977; PubMed=12381787;  
Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,  
Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,  
Priemore R.D., Arigoni F.,  
"The genome sequence of Bifidobacterium longum reflects its adaptation  
to the human gastrointestinal tract."  
Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
EMBL; AB014793; AAN25367.1; -  
Hypothetical protein; Complete proteome.  
SEQUENCE 343 AA; 37507 MW; 8AB35FF32D22E059 CRC64;

Query Match 37.3%; Score 57.5; DB 16; Length 343;  
Best Local Similarity 54.2%; Pred. No. 9.3;  
Matches 13; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

9 RGR-----ASWCWALALMLAVP 27  
266 RGRCSYVAASWVWVALLILAAVP 289

SUITE 6  
4X72  
084X72 PRELIMINARY; PRT; 366 AA.

01-JUN-2003 (TREMBlrel. 24, Created)  
01-JUN-2003 (TREMBlrel. 24, last sequence update)  
01-JUN-2003 (TREMBlrel. 24, last annotation update)  
CR066 protein.  
Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
NCBI\_TaxID=3055;  
[1]  
SEQUENCE FROM N.A.  
Li J.B., Lin S., Jia H., Wu H., Roe B.A., Kulip D., Stormo G.D.,  
Dutcher S.K.,  
"Finished genomic sequence in Chlamydomonas."  
Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.  
EMBL; AY207499; AAO32622.1; -  
SEQUENCE 366 AA; 38816 MW; 99F528BB9E625BB CRC64;

Query Match 37.0%; Score 57; DB 10; Length 366;  
Best Local Similarity 45.5%; Pred. No. 11;  
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

4 GAGAGRGRASGCMALALMLAV 25  
315 GGGGQOPRRRCWAEGLMVEV 336

SUITE 7  
9464  
069464 PRELIMINARY; PRT; 214 AA.

01-AUG-1998 (TREMBlrel. 07, Created)  
01-AUG-1998 (TREMBlrel. 07, last sequence update)  
01-JUN-2003 (TREMBlrel. 24, last annotation update)  
Putative integral membrane protein (conserved integral membrane  
protein).  
ML1666 OR MLCB1243.07.

OS Mycobacterium leprae.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrall B.G.,  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
DR EMBL; AL023635; CAA19188.1; -  
DR EMBL; AL583923; CAC30619.1; -  
DR PIR; T44701; T44701.  
DR Leproma; ML1666; -  
KM Complete proteome.  
SQ SEQUENCE 214 AA; 23402 MW; 5B48414709828FA0 CRC64;

Query Match 36.4%; Score 56; DB 16; Length 214;  
Best Local Similarity 41.7%; Pred. No. 9.5;  
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 3 RGAGARGRASGCMALALMLAV 26  
DB 175 RTSALRGTYQRMSTATLWFTV 198

RESULT 8  
088SV3 PRELIMINARY; PRT; 485 AA.

01-JUN-2003 (TREMBlrel. 24, Created)  
01-JUN-2003 (TREMBlrel. 24, last sequence update)  
01-OCT-2003 (TREMBlrel. 25, last annotation update)  
DE Cardiolipin synthetase 2 (EC 2.7.8.-).  
GN C15 OR IP 3273.  
OS Lactobacillus plantarum.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1590;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTMB 8826 / WCFS1;  
RX MEDLINE=22480296; PubMed=12566566;  
RA Kleerebezem M., Boekhorst J., van Krenenburg R., Molenaar D.,  
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,  
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
RA De Vos W.M., Siezen R.J.,  
RT "Complete genome sequence of Lactobacillus plantarum WCFS1."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).  
DR EMBL; AL935261; CAD65406.1; -  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR001736; P1D.  
DR Pfam; PF00614; PLDC; 2.  
DR PROSITE; PS50035; PLD; 2.  
KW Transferase; Complete proteome.

SEQUENCE 485 AA; 55778 MW; C725C28841DF8D31 CRC64;

Query Match 36.4%; Score 56; DB 16; Length 485;  
Best Local Similarity 46.7%; Pred. No. 20;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 12 ASWCWALALMLAV 26

36 ATWAMLLVLTLPRI 50

RESULT 9

12XSYO PRELIMINARY; PRT; 412 AA.

01-OCT-2003 (TREMBlrel. 25, Created)  
01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
OSUNB0056F09.16 protein.

OSUNB0056F09.16.

Oryza sativa (Rice).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

NCBI\_TaxID=4530;

SEQUENCE FROM N.A.

Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,

Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.T.,

Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,

Shang H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,

Huang Y.C., Li Y., Zhu J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,

Chen L., Fan D.L., Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,

Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,

Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,

Zhang R.Q., Guan J.P., Hong G.F.,

Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

EMBL: AL606602; CAE01753.1; -

SEQUENCE 412 AA; 45803 MW; 1A8FE532333606C8 CRC64;

Query Match 36.0%; Score 55.5; DB 10; Length 412;

Best Local Similarity 40.7%; Pred. No. 20;

Matches 11; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

4 GAGAGAG-----RASWCALALMLAV 25

19 GSGGSGGLPRATRKEMWMSGILLKAV 45

RESULT 10

8XWMA PRELIMINARY; PRT; 695 AA.

01-MAR-2002 (TREMBlrel. 20, Created)

01-OCT-2003 (TREMBlrel. 25, Last sequence update)

Probable bifunctional: uroporphyrin-III C-methyltransferase and

uroporphyrinogen-III synthase transmembrane protein

(EC 2.1.1.107).

RSC2356 OR RS01190.

Ralstonia solanacearum (Pseudomonas solanacearum).

Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

Burkholderiaceae; Ralstonia.

NCBI\_TaxID=305;

SEQUENCE FROM N.A.

STRAIN=GM11000.

MEDLINE=21681879; PubMed=11823852;

Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

Arlat M., Billault A., Broclet P., Camus J.C., Catolico L.,

Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,

Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,

Wiesendach J., Theault P., Whalen M., Winkler P., Levy M.,

"Genome sequence of the plant pathogen Ralstonia solanacearum";

Nature 415:497-502(2002).

EMBL: AL646069; CAD16063.1; -

GO: 0008168; F:methyltransferase activity; IEA.

GO: 0016740; F:transferase activity; IEA.

GO: 0004851; F:uroporphyrin-III C-methyltransferase activity; IEA.

DR GO: 0004852; F:uroporphyrinogen-III synthase activity; IEA.

DR InterPro: IPR007470; DUF513.

DR InterPro: IPR003754; HEM4\_synth.

DR Pfam: PF02602; HEM4; 1.

DR Pfam: PF04375; HemX; 1.

KM Transferase; Methyltransferase; Complete proteome.

SEQUENCE 695 AA; 73670 MW; 1649C5BA781DF05 CRC64;

Query Match 36.0%; Score 55.5; DB 16; Length 695;

Best Local Similarity 52.0%; Pred. No. 32;

Matches 13; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

5 AGAARGASW-CALMLMLAVPG 28

344 AGAGRGGLWMLALMLAVVAG 368

RESULT 11

07X917 PRELIMINARY; PRT; 147 AA.

01-OCT-2003 (TREMBlrel. 25, Created)

01-OCT-2003 (TREMBlrel. 25, Last sequence update)

01-OCT-2003 (TREMBlrel. 25, Last annotation update)

OSUNB0040D17.2 protein (OSUNB0033G08.19 protein).

OSUNB0040D17.2 OR OSUNB0033G08.19.

Oryza sativa (Rice).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

NCBI\_TaxID=4530;

SEQUENCE FROM N.A.

Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,

Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Q.J., Zhang L.,

Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,

Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,

Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Tu Y.F., Jia J., Yin H.F.,

Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,

Zhang Y., Cai Z., Kang H., Chen X.Y., Shao C.Y., Sun Y.,

Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,

Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.,

Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,

Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,

Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,

Shang H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,

Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,

Chen L., Fan D.L., Meng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,

Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,

Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,

Zhang R.Q., Guan J.P., Hong G.F.,

Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

EMBL: AL662965; CAD19623.1; -

EMBL: AL606989; CAE02003.1; -

SEQUENCE 147 AA; 14767 MW; 024314375754B785 CRC64;

Query Match 35.7%; Score 55; DB 10; Length 147;

Best Local Similarity 33.3%; Pred. No. 9.2;

Matches 9; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

2 FRGAGARGASW-CALMLMLAVVPG 28

99 FRACSSRRDGMWMSCVMAQITPG 125

RESULT 12

07WFK8 PRELIMINARY; PRT; 304 AA.

07WFK8



PROTEIN C AND MRNA (BY SIMILARITY).

EMBL: D84252; BA332664.1; -.  
HSSP: S27958; IAIIV.  
MEROPS; U39.001; -.  
GO; GO:0016021; C:integral to membrane; IEA.  
GO; GO:0019028; C:viral capsid; IEA.  
GO; GO:0019031; C:viral envelope; IEA.  
GO; GO:0005524; F:ATP binding; IEA.  
GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
GO; GO:0003723; F:RNA binding; IEA.  
GO; GO:0009368; F:serine-type peptidase activity; IEA.  
GO; GO:0005198; F:structural molecule activity; IEA.  
GO; GO:0016740; F:transferase activity; IEA.  
GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
GO; GO:0006350; E:transcription; IEA.  
GO; GO:0019079; E:viral genome replication; IEA.  
GO; GO:0019087; E:viral transformation; IEA.  
InterPro; IPR009003; Cys\_Ser\_trypsin.  
InterPro; IPR001410; DEAD.  
InterPro; IPR002522; HCV\_capsid.  
InterPro; IPR002521; HCV\_core.  
InterPro; IPR002519; HCV\_env.  
InterPro; IPR002531; HCV\_NS1.  
InterPro; IPR002518; HCV\_NS2.  
InterPro; IPR000745; HCV\_NS4.  
InterPro; IPR001490; HCV\_NS4b.  
InterPro; IPR002868; HCV\_NS5a.  
InterPro; IPR002166; HCV\_NS5a.  
InterPro; IPR001650; Helicase\_C.  
InterPro; IPR003006; IG\_MHC.  
InterPro; IPR004109; peptidase\_C9.  
InterPro; IPR007095; RNA\_pol\_DS\_PS.  
InterPro; IPR007094; RNA\_pol\_PSVir.  
Pfam; PF01543; HCV\_capsid; 1.  
Pfam; PF01542; HCV\_core; 1.  
Pfam; PF01539; HCV\_env; 1.  
Pfam; PF01560; HCV\_NS1; 1.  
Pfam; PF01538; HCV\_NS2; 1.  
Pfam; PF02907; HCV\_NS3; 1.  
Pfam; PF01006; HCV\_NS4a; 1.  
Pfam; PF01001; HCV\_NS4b; 1.  
Pfam; PF01506; HCV\_NS5a; 1.  
Pfam; PF00271; helicase\_C; 1.  
Pfam; PF00998; Viral\_RdRp; 1.  
ProDom; PD186062; HCV\_NS1; 1.  
SMART; SM00487; DEXDC; 1.  
PROSITE; PS00290; IG\_MHC; 1.  
Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.

SEQUENCE 3019 AA; 328230 MW; E26750E07BCBC310 CRC64;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: June 8, 2004, 13:55:43 ; Search time 7.95833 Seconds

(without alignments)  
1065.101 Million cell updates/sec

file: US-10-010-050a-2\_COPY\_1\_30

effect score: 169

sequence: 1 MRRGAGAAAGRAAGCAATLALTLAVVPGWS 30

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

initium DB seq length: 0  
maximum DB seq length: 2000000000

opt-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :

A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	169	100.0	347	2	AAV41323 Human sec
2	169	100.0	358	4	AAW93870 Human pol
3	162	95.9	346	2	AAW92967 Human zsi
4	60.5	35.8	192	4	AAU29418 Human G p
5	60.5	35.8	192	5	ABG60706 Novel G p
6	59.5	35.2	127	4	AAU14162 Human nov
7	58	34.3	73	4	ABG29252 Novel hum
8	58	34.3	188	5	ABU60973 Lung spec
9	57.5	34.0	19938	6	ABP76682 Streptomy
10	56	33.1	214	5	ABU05547 M. tuberc
11	56	33.1	403	5	ABU65190 Human NOV
12	56	33.1	1093	2	AAW41001 Human myo
13	55.5	32.8	371	4	AAW93279 Human pro
14	55.5	32.8	371	5	AAO18510 Human ins
15	55.5	32.8	371	5	ABG32373 Human ins
16	55.5	32.8	421	6	ADA10702 Human PHB
17	54.5	32.2	169	2	AAV59780 Human PHB
18	54.5	32.2	169	2	AAV59780 Human PHB
19	54	32.0	78	4	ABV60506 Human nov
20	53.5	31.7	95	6	ABV58357 Human ner
21	53.5	31.7	117	4	AAU54332 Bcl1040 p
22	53.5	31.7	117	6	AAU54332 Propionib
23	53	31.4	34	2	AAV43189 Thiredox
24	53	31.4	65	2	AAV00765 Pylon pro
25	53	31.4	117	6	ADA54650 Human pro

26	53	31.4	211	4	ABG01709	ABG01709 Novel hum
27	53	31.4	243	4	AAU46489	AAU46489 Propionib
28	53	31.4	243	6	ABW43008	ABW43008 Propionib
29	53	31.4	346	5	ABG96249	ABG96249 Maize per
30	53	31.4	532	3	AAV84593	AAV84593 Amino aci
31	53	31.4	600	4	ABG07764	ABG07764 Novel hum
32	52.5	31.1	199	4	AAU28370	AAU28370 Novel hum
33	52.5	31.1	199	7	ADSO9107	ADSO9107 Novel pro
34	52.5	31.1	199	7	ADSO9108	ADSO9108 Novel pro
35	52.5	31.1	206	3	AAW42500	AAW42500 Human ORF
36	52.5	31.1	244	4	AAW42195	AAW42195 Human ORF
37	52.5	31.1	267	4	AAW9568	AAW9568 Human pol
38	52.5	31.1	333	4	AAW91017	AAW91017 C glutam
39	52.5	31.1	341	3	AAV69880	AAV69880 B. lactof
40	52.5	31.1	360	6	ABU22472	ABU22472 Protein e
41	52.5	31.1	442	1	AAV90475	AAV90475 N-termina
42	52.5	31.1	551	2	AAW6503	AAW6503 Cattle MI
43	52.5	31.1	575	1	AAV70195	AAV70195 Sequence
44	52.5	31.1	575	1	AAV90547	AAV90547 Bovine Mt
45	52.5	31.1	575	2	AAV76500	AAV76500 Cattle MI

## ALIGNMENTS

RESULT 1	AAV41323	AAV41323 standard; protein, 347 AA.
XX	AAV41323;	
XX	02-DEC-1999	(first entry)
DE	Human secreted protein encoded by gene 16 clone HMZAD77.	
XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;	
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;	
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;	
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;	
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
OS	Homo sapiens.	
XX	W09947540-A1.	
XX	23-SEP-1999.	
PD	18-MAR-1999;	99WO-US005804.
XX	19-MAR-1998;	98US-0078563P.
XX	19-MAR-1998;	98US-0078566P.
XX	19-MAR-1998;	98US-0078573P.
XX	19-MAR-1998;	98US-0078574P.
XX	19-MAR-1998;	98US-0078577P.
XX	19-MAR-1998;	98US-0078578P.
XX	19-MAR-1998;	98US-0078579P.
XX	19-MAR-1998;	98US-0078581P.
XX	01-APR-1998;	98US-0080312P.
XX	01-APR-1998;	98US-0080313P.
XX	01-APR-1998;	98US-0080314P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;	
XX	Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;	
XX	Shi Y, Moore PA,	
XX	WPI, 1999-562050/47.	
XX	N-PSDB; AA224826.	







Sequence encoding a portion of a G protein coupled receptor (ngPCR-x).  
(1) is used to produce a recombinant ngPCR-x polypeptide. A polypeptide encoded by (1) is used to induce an immune response in a mammal. ngPCR-x is used to identify a compound that binds to it and/or modulates it's activity. (1) is used to identify animal homologues of ngPCR-x. (1) can be used to diagnose a human subject as having a brain or genetic predisposition disorder, such as a mental disorder. (1) is used to screen for an ngPCR-x related disorder including thyroid disorders (e.g. thyrotoxicosis, myxoedema), renal failure, inflammatory conditions (e.g. Crohn's disease), diseases related to cell differentiation and homeostasis, rheumatoid arthritis, autoimmune disorders, movement disorders, CNS disorders, viral infections (e.g. Human immunodeficiency virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity, anorexia, cardiomyopathies), proliferative diseases and cancers (e.g. psoriasis, lung cancer), hormonal disorders, sexual dysfunction and hereditary mental disorders in a human patient. A host cell comprising (1) is used to screen for a modulator of ngPCR-x activity. ngPCR-x is used to identify compounds that can treat mental disorders. The polypeptide encoded by (1) is used to purify a G protein from a sample. This is the amino acid sequence of a novel G protein coupled receptor (ngPCR-x) protein described in the invention

Sequence 192 AA;

Query Match 35.8%; Score 60.5; DB 5; Length 192;  
Best Local Similarity 39.3%; Pred. No. 2.7;  
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

4 GAGAGRGRASWCW-----ALALIM 22  
35 GKAVRGRPCWCMPCQPALVITATLV 62

RESULT 6  
AAU14162  
D AAU14162 standard; protein; 127 AA.

AAU14162;  
24-OCT-2001 (first entry)

Human novel protein #33.  
Human; novel protein; Antianemic; antiinflammatory; immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic; anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral; antibacterial; antiallergic; dermatological; haemostatic; antisthmatic; thrombolytic; immunogen; antibody; gene therapy; neurological disorder; Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis; tissue regeneration; immune disorder.  
Homo sapiens.

WO200155437-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US002623.

25-JAN-2000; 2000US-00491404.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-451939/48.

N-PSDB; AAS22467.

Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage.  
Example 4; Page 543-544; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. CC Polynucleotides of the invention are used as probes and primers, for CC sequencing, for chromosome or gene mapping, in the production of CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene CC therapy. Polypeptides of the invention can be used to target drugs to a CC tumour, in assays to determine biological activity, to raise CC antibodies/ elicit an immune response, to determine quantitative protein CC levels, as tissue markers, and to isolate receptors or ligands. CC Polypeptides of the invention may also be useful in treating platelet CC disorders, stem cell disorders, regenerating bone, cartilage, tendon, CC ligament and/or nerve tissue, wound healing, treating burns, promoting CC the proliferation, differentiation and survival of stem cells, as a CC contraceptive, treating osteoporosis and osteoarthritis, anaemia, CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft- CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory CC diseases, nervous system disorders, and infection. The present sequence CC represents a protein of the invention

Sequence 127 AA;

Query Match 35.2%; Score 59.5; DB 4; Length 127;  
Best Local Similarity 40.0%; Pred. No. 2.4;  
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

3 RGAGAAAGR-----ASWCALALL---MLAVVPG 28  
25 KRGERTGRVTTTKLQWMLGLDSTWVALTTG 59

RESULT 7  
ABG29252  
ID ABG29252 standard; protein; 73 AA.

ABG29252;  
18-FEB-2002 (first entry)

Novel human diagnostic protein #29243.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS93439.

New isolated polynucleotide and encoded polypeptides, useful in PT diagnosis, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity.



2905 RRAAVTPXGARGRVWCGPVGSGWICAPP 2933

RESULT 10  
ABU05547  
ID ABU05547 standard; protein, 214 AA.  
KC ABU05547;  
KT 08-APR-2003 (first entry)  
KE M. tuberculosis and M. leprae marker protein #198.  
KW Mycobacteriosis; survival; virulence; protective antigen; vaccine;  
KW mycobacterial disease; tuberculosis; leprosy.  
KS Mycobacterium tuberculosis.  
KN Mycobacterium leprae.  
KO W0200274903-A2.  
KD 26-SEP-2002.  
KP 22-FEB-2002; 2002MO-IB001973.  
KR 22-FEB-2001; 2001US-0270123P.  
KS (INSP ) INST PASTEUR.  
KT Cole S;  
KW WPI; 2002-759885/82.  
KT Identifying and selecting genes for survival or virulence of mycobacteria  
KT by a comparative genomic analysis of the sequences of Mycobacterium  
KT tuberculosis and M. leprae.  
KS Claim 17; Page 376-377; 874pp; English.  
KC This invention relates to a novel method for identifying essential genes  
KC for survival or virulence of mycobacteria species. The method comprises  
KC aligning the genomic sequence of a first mycobacterium species on a  
KC genomic sequence of a second mycobacterium species and selecting a  
KC polynucleotide sequence that is highly conserved in both genomes with no  
KC counterparts in other bacterial genomic sequences and that corresponds to  
KC an essential gene for the survival or virulence of mycobacterium species.  
KC The method of the invention is useful for detecting M. tuberculosis or M.  
KC leprae infection. The method reduces the number of potential new targets  
KC and protective antigens for new drugs and vaccine compositions to treat  
KC and prevent mycobacterial diseases, particularly tuberculosis and  
KC leprosy. The present sequence represents a marker protein from  
KC Mycobacterium tuberculosis and Mycobacterium leprae identified using the  
KC method of the invention  
KQ Sequence 214 AA;  
Query Match 33.1%; Score 56; DB 5; Length 214;  
Best Local Similarity 41.7%; Pred. No. 12;  
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

3 RGAGARGRASMCMALLMLLAAYV 26  
175 RTSIARGIYQWRWSITLWETIV 198

RESULT 11  
ABU65190  
ID ABU65190 standard; protein; 403 AA.  
KC ABU65190;  
KT 20-MAY-2003 (first entry)

DE Human NOV105a protein.  
XX  
XX NOVX; cytosolic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;  
KW hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;  
KW human.  
XX  
XX Homo sapiens.  
OS  
EN W0200272757-A2.  
PN  
XX 19-SEP-2002.  
PD  
XX  
PF 08-MAR-2002; 2002MO-US006908.  
PE  
XX 08-MAR-2001; 2001US-0274101P.  
PR 08-MAR-2001; 2001US-0274194P.  
PR 08-MAR-2001; 2001US-0274281P.  
PR 08-MAR-2001; 2001US-0274322P.  
PR 09-MAR-2001; 2001US-0274849P.  
PR 12-MAR-2001; 2001US-0275235P.  
PR 13-MAR-2001; 2001US-0275578P.  
PR 13-MAR-2001; 2001US-0275579P.  
PR 13-MAR-2001; 2001US-0275579P.  
PR 13-MAR-2001; 2001US-0275601P.  
PR 14-MAR-2001; 2001US-0276000P.  
PR 16-MAR-2001; 2001US-0276776P.  
PR 19-MAR-2001; 2001US-0276994P.  
PR 20-MAR-2001; 2001US-0277239P.  
PR 20-MAR-2001; 2001US-0277321P.  
PR 20-MAR-2001; 2001US-0277327P.  
PR 21-MAR-2001; 2001US-0277791P.  
PR 22-MAR-2001; 2001US-0277833P.  
PR 23-MAR-2001; 2001US-0278152P.  
PR 26-MAR-2001; 2001US-0278894P.  
PR 27-MAR-2001; 2001US-0278999P.  
PR 27-MAR-2001; 2001US-0279036P.  
PR 28-MAR-2001; 2001US-0279344P.  
PR 30-MAR-2001; 2001US-0277338P.  
PR 30-MAR-2001; 2001US-0279995P.  
PR 30-MAR-2001; 2001US-0280233P.  
PR 02-APR-2001; 2001US-0280802P.  
PR 02-APR-2001; 2001US-0280822P.  
PR 02-APR-2001; 2001US-0280900P.  
PR 04-APR-2001; 2001US-0281194P.  
PR 13-APR-2001; 2001US-0283675P.  
PR 30-APR-2001; 2001US-0287424P.  
PR 02-MAY-2001; 2001US-0288066P.  
PR 03-MAY-2001; 2001US-0288342P.  
PR 03-MAY-2001; 2001US-0288528P.  
PR 15-MAY-2001; 2001US-0291190P.  
PR 16-MAY-2001; 2001US-0291099P.  
PR 16-MAY-2001; 2001US-0291240P.  
PR 30-MAY-2001; 2001US-0294485P.  
PR 31-MAY-2001; 2001US-0294889P.  
PR 31-MAY-2001; 2001US-0294899P.  
PR 18-JUN-2001; 2001US-0299027P.  
PR 19-JUN-2001; 2001US-0299303P.  
PR 19-JUN-2001; 2001US-0299310P.  
PR 10-JUL-2001; 2001US-0304354P.  
PR 31-JUL-2001; 2001US-0309198P.  
PR 16-AUG-2001; 2001US-0312903P.  
PR 10-SEP-2001; 2001US-0318462P.  
PR 12-SEP-2001; 2001US-0318770P.  
PR 27-SEP-2001; 2001US-0325430P.  
PR 27-SEP-2001; 2001US-0325681P.  
PR 18-OCT-2001; 2001US-0330380P.  
PR 31-OCT-2001; 2001US-0335301P.  
PR 14-NOV-2001; 2001US-0332172P.  
PR 14-NOV-2001; 2001US-0332271P.  
PR 14-NOV-2001; 2001US-0332272P.  
PR 14-NOV-2001; 2001US-0333184P.  
PR 14-NOV-2001; 2001US-0333722P.  
PR 21-NOV-2001; 2001US-0332094P.  
PR 03-DEC-2001; 2001US-0337426P.

R 03-DEC-2001; 2001US-0338092P.  
R 04-DEC-2001; 2001US-0337185P.  
R 03-JAN-2002; 2002US-0345705P.  
R 07-MAR-2002; 2002US-00092900.  
X  
X (CURA-) CURAGEN CORP.  
X  
X Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L,  
I Zehrhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R,  
I Pattunajan M, Gangoli E, Vernet CAM, Guo X, Tcheurev V,  
I Fernandes ER, Casman SJ, Malayankar UM, Gerlach V, Liu Y, Anderson D,  
I Spaderna SK, Catterton E, Burgess C, Lette M, Zhong H, Alsbrook UB,  
I Lepley DM, Rieger DK;  
X WPI; 2002-723332/78.  
R N-PSDB; ABX97157.  
R  
R NOVX polypeptides and polynucleotides, useful for preventing or treating  
T a disorder associated with aberrant NOVX expression or activity e.g.,  
T cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial  
T asthma.  
X  
X Claim 1; Page 410; 1103pp; English.  
X  
X This invention describes novel human NOVX polypeptides which have  
C cytoskeletal, cardiac, antiarteriosclerotic, antiasthmatic and hypotensive  
C activity. Pharmaceutical compositions comprising the NOVX proteins or  
C nucleic acid molecules or NOVX antibodies are useful for preventing or  
C treating a disorder associated with aberrant NOVX expression or activity  
C e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial  
C asthma. The products of the invention can be used for gene therapy or in  
C a vaccine. ABU65041-ABU65218 represent the NOVX polypeptides encoded by  
C AX97008-ABX97185  
X  
Q Sequence 403 AA:  
  
Query Match 33.1%; Score 56; DB 5; Length 403;  
Best Local Similarity 37.8%; Pred. No. 24;  
Matches 14; Conservative 1; Mismatches 14; Indels 8; Gaps 1;  
  
Y 2 RRGAGAGRGASWCALALIMLAVVPG-----WS 30  
b 118 RRGISLRGAAGLAWRLFLIMLAVLPSIACILYYWS 154  
  
RESULT 12  
D AAR41001 standard; protein; 1093 AA.  
X  
C AAR41001;  
X  
X 25-MAR-2003 (revised)  
T 25-FEB-1994 (first entry)  
X  
X Human myotonic dystrophy gene protein.  
X  
X Abnormality: muscular dystrophy; CHR 19; chromosome 19; protein kinase;  
W polymerase chain reaction; brain.  
X  
X Homo sapiens.  
X  
X Key Location/Qualifiers  
H Region 1..1093  
T  
T sequence "encoded by predicted reading frame a, x's in the  
T notes indicate stop codons in the reading frame"  
X  
X WC9317104-A1.  
X  
X 02-SEP-1993.  
X  
X 19-FEB-1993; 93MO-US001545.  
X  
X 20-FEB-1992; 92US-00839255.  
X

XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
XX Brook JD, Housman DE;  
PI  
XX WPI; 1993-288410/36.  
DR  
XX  
XX  
PT DNA sequence of myotonic dystrophy gene - used to produce probes and  
PT identify CHR 19 abnormality and protein kinase responsible.  
XX  
XX  
PS Disclosure; Fig 6; 64pp; English.  
XX  
XX The sequence is that encoded by predicted reading frame a of the human  
CC myotonic dystrophy (DM) gene. It may be used in the identification of  
CC individuals affected by DM. (updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1093 AA;  
  
Query Match 33.1%; Score 56; DB 2; Length 1093;  
Best Local Similarity 39.5%; Pred. No. 67;  
Matches 15; Conservative 1; Mismatches 6; Indels 16; Gaps 3;  
  
QY 3 RCGAGAA-----RGRASWCALALIMLAVVPGWS 30  
Db 207 RGHGAACPTCQPRCGXGSSSWCWTTRA-SW-----GWS 238  
  
RESULT 13  
ID AAG93279 standard; protein; 371 AA.  
XX  
XX AAG93279;  
AC  
XX  
DT 13-SEP-2001 (first entry)  
XX  
XX Human protein HP03145.  
DE  
XX  
XX Human; gene therapy; tumour.  
KM  
XX  
OS Homo sapiens.  
XX  
XX WO200142302-A1.  
PN  
XX  
PD 14-JUN-2001.  
XX  
XX  
PF 06-DEC-2000; 2000WO-JP008631.  
XX  
XX 06-DEC-1999; 99JP-00346863.  
PR 06-DEC-1999; 99JP-00346864.  
PR 08-FEB-2000; 2000JP-00031062.  
PR 10-FEB-2000; 2000JP-00034090.  
PR 10-FEB-2000; 2000JP-00034091.  
PR 14-FEB-2000; 2000JP-00035829.  
PR 14-FEB-2000; 2000JP-00035829.  
PR 14-MAR-2000; 2000JP-00071161.  
PR 30-MAY-2000; 2000JP-00160851.  
XX  
XX  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX  
XX Kato S, Eguchi C, Saeki M;  
PI  
XX  
XX  
DR WPI; 2001-381646/40.  
DR N-PSDB; AAH68564.  
XX  
XX Human protein originated from tumor cell line, applicable as drug,  
PT reagent for studying intracellular protein networks and protein source  
PT for drug screening, also encoded cDNA for gene diagnosis and gene  
PT therapy.  
XX  
XX  
PS Claim 1; Page 246-248; 471pp; Japanese.  
XX  
XX The present sequence is a human protein. The human protein, preferably  
CC originated from tumour cell line, is applicable as a drug, a reagent for

studying intracellular protein networks and a protein source for screening proteins for binding low molecular weight drugs. The human protein coding sequence is useful for gene diagnosis and gene therapy, expression vectors and transformant cells for detection of ligands and receptors

50 Sequence 371 AA;

Query Match	32.8%	Score 55.5;	DB 4;	Length 371;
Best Local Similarity	55.6%;	Pred. No. 26;		
Matches 15; Conservative	0;	Mismatches 5;	Indels 7;	Gaps 2;

2y 3 RGAGARGRASWCWALATLTLAVVPGW 29  
| | | | | | | | | |  
2b 5 RAAGFARGLR----ALATLTL----PGW 24

RESULT 14	
AA018510	
ID	AA018510 standard; protein; 371 AA.

AC AA018510;  
KX  
DT 11-OCT-2002 (First entry)

Human insulin receptor signaling modifier SEQ ID NO: 42.

KM Human; insulin receptor signaling; insulin receptor signaling modifier;  
KM ISM; diabetes; metabolic syndrome; antidiabetic.

25 Homo sapiens.

W0200255664-A2.

18-JUL-2002

11-JAN-2002; 2002WO-US001048.

2R	12-JUN-2001	2001US-02612266
2R	12-JUN-2001	2001US-02613033
2R	12-JUN-2001	2001US-02613040
2R	12-JUN-2001	2001US-02613355
2R	12-JUN-2001	2001US-02613366
2R	12-JUN-2001	2001US-02613616
2R	12-JUN-2001	2001US-02614566
2R	12-JUN-2001	2001US-02614570
2R	12-JUN-2001	2001US-02614582
2R	12-JUN-2001	2001US-02614599
2R	12-JUN-2001	2001US-02614616
2R	12-JUN-2001	2001US-02615186
2R	12-JUN-2001	2001US-02615316
2R	12-JUN-2001	2001US-02615322
2R	12-JUN-2001	2001US-02615899
2R	12-JUN-2001	2001US-02615906
2R	12-JUN-2001	2001US-02616946
2R	12-JUN-2001	2001US-02616956
2R	12-JUN-2001	2001US-02616976

2A (EXEL-) EXELIXIS INC

21 Seidel-Dugan C, Ferguson KC, Kidd T;

DR WPI; 2002-599664/64.

NR N-PSDB; AAL48629.

Identifying an insulin receptor signaling modulator, useful as drug targets for treating diabetes or metabolic disorders, comprises contacting an assay system comprising insulin receptor signaling modifiers with a test agent.

25 Disclosure; Page 134-135; 232pp; English.

23 The present invention relates to a method of identifying a candidate

insulin receptor (INR) signaling modulating agent, involving contacting an assay system comprising an insulin receptor signaling modifier (ISM) polypeptide or nucleic acid with a test agent, and detecting a test agent biased activity of the assay system. The method is useful for identifying candidate INR signaling modulating agents. ISM genes may be used as drug targets for treatment of disorders related to INR signaling such as diabetes or metabolic syndrome. ISM nucleic acids and polypeptides are useful for identifying and testing agents that modulate ISM function and for other applications related to the involvement of ISM in INR signaling, and for identifying subjects having a predisposition to such diseases associated with INR signaling. The present sequence is an ISM protein described in the exemplification of the invention

Sequence 371 AA;

Query Match	32.8%	Score	55.5	DB	5	Length	371
Best Local Similarity	55.6%	Pred. No.	26				
Matches	15	Conservative	0	Mismatches	5	Indels	7
						Gaps	2

OY            3 RGAGAA<sup>-</sup>RGRASWC<sup>-</sup>MALAL<sup>-</sup>LWLAVPGW 25  
               | | | | | | | | | |  
DB            5 RAAGFARGLR---ATLAWL---PGW 24

RESULT 15  
ABG32373  
ID ABG32373 standard; protein; 371 AA

AC ABG32373-

DT 15-NOV-2002 (first entry)

DE Novel human enzyme, NZMS-3, Incyte 2900469CD1.

KM Human; NZMS-3; immune system disorder; AIDS; contact dermatitis;  
 KM acquired immune deficiency syndrome; Chediak-Higashi syndrome;  
 KM common variable immunodeficiency; developmental disorder; catara-  
 KM Cushing's syndrome; hyperthyroidism; eye disorder; glaucoma; asch  
 KM metabolic disorder; cystic fibrosis; sickle-cell anaemia; epilepsy  
 KM smooth muscle disorder; lactic acidosis; neurological disorder;  
 KM Parkinson's disease; cardiac disorder; congestive heart failure;  
 KM myocardial infarction; pulmonary disorder; emphysema; bronchitis;  
 KM parasitic infection; amoebiasis; filariasis; cancer; leukaemia;  
 KM cell proliferative disorder; Inocyte 2900469CD1, enzyme

OS Homo sapiens.

PN WO200264795-A2.

PD 22-AUG-2002

PF 08-FEB-2002; 2002WO-US003814.

PR	09-FEB-2001	2001US-0268113P
PR	15-FEB-2001	2001US-0269215P
PR	27-FEB-2001	2001US-0272271P
PR	07-MAR-2001	2001US-0274091P
PR	09-MAR-2001	2001US-0274423P
PR	23-MAR-2001	2001US-0278479P
PR	23-MAR-2001	2001US-0278480P

PA (INCY-) INCYTE GENOMICS INC.

PI Sanjmwala MM, Lu Y, Lee EA, Hafalia AJL, Warren BA, Baughn MR  
PI tang TY, Yue H, Yao MG, Lee S, Thornton M, Walla NK, Xu Y;  
PI Tran UK, Lal PG, Lu DM, Swarnaker A, Jones KA, Rung HZ;  
XX  
WPI, 2002-643488/69.  
DR N-PSDB; ABK91314.

DR N-PSDB; ABK91314.

PT New isolated human enzymes (NZMS), useful for diagnosing, treating and  
PT preventing immune system, developmental, eye, metabolic, smooth muscle  
PT neurological and cardiac disorders, cancer, or parasitic

infections.

Claim 1; Page 149-150; 170pp; English.

The invention relates to a new isolated human enzyme (NZMS) (I) and the encoding nucleic acid (II). The polypeptides and polynucleotides are useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of NZMS. The NZMS or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds that specifically bind to or modulate the activity of the polypeptide. The antibody is useful in diagnosing a condition or disease associated with the decreased expression or overexpression of NZMS and in detecting (I). (I) is useful in the production of antibodies. The microarray is useful in monitoring or measuring gene expression profiles. The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of NZMS, such as immune system disorders (e.g. acquired immune deficiency syndrome (AIDS), contact dermatitis), immune deficiencies (e.g. common variable immunodeficiency, Chediak-Higashi syndrome), developmental disorders (e.g. Cushing's syndrome, hypothyroidism), eye disorders (e.g. glaucoma, cataract), metabolic disorders (e.g. cystic fibrosis, sickle-cell anaemia), smooth muscle disorders (e.g. asthma, lactic acidosis), neurological disorders (e.g. Parkinson's disease, epilepsy), cardiac disorders (e.g. congestive heart failure, myocardial infarction), pulmonary disorders (emphysema, bronchitis), parasitic infections (amoebiasis, filariasis), and cell proliferative disorders (e.g. cancer, leukaemia). The microarray is useful in monitoring or measuring protein-protein interactions, drug-target interactions, and gene expression profiles. The present sequence represents the amino acid sequence of one of 11 novel human enzymes (NZMS1-11) of the invention

Sequence 371 AA;

Query Match 32.8%; Score 55.5; DB 5; Length 371;  
Best Local Similarity 55.6%; Pred. No. 26;  
Matches 15; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

3 RGAGARGASWCMALIMLAVPGW 29  
5 RAGFARGLR--ALALAWL--PGW 24

Search completed: June 8, 2004, 14:03:57  
Job time : 8.95833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 14:00:33 ; Search time 2.3333 Seconds  
(without alignments)  
663.762 Million cell updates/sec

Title: US-10-010-050A-2\_COPY\_1\_30  
Effect score: 169  
Sequence: 1 MRRGAGAGRGASWCALALLMLAVPGWS 30

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

otal number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents.AA:  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiller.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	32.5	109	4	US-09-252-991A-31463 Sequence 31463, A
2	54.5	32.2	263	4	US-09-252-991A-16906 Sequence 16906, A
3	53	31.4	74	4	US-09-621-976-6090 Sequence 6090, A
4	53	31.4	215	3	US-09-220-528-104 Sequence 104, App
5	53	31.4	228	4	US-09-252-991A-32898 Sequence 32898, A
6	52.5	31.1	575	1	US-07-683-957B-2 Sequence 2, App1
7	52	30.8	170	4	US-09-252-991A-1705 Sequence 16705, A
8	52	30.8	217	4	US-09-252-991A-1862 Sequence 1862, A
9	51.5	30.5	102	4	US-09-252-991A-23296 Sequence 23296, A
10	51.5	30.5	194	4	US-09-252-991A-21199 Sequence 21199, A
11	51.5	30.5	370	4	US-09-252-991A-27810 Sequence 27810, A
12	51	30.2	85	4	US-09-431-705-22 Sequence 22, App1
13	51	30.2	88	4	US-09-205-258-274 Sequence 274, App
14	51	30.2	1310	4	US-09-170-496D-290 Sequence 290, App
15	51	30.2	1310	4	US-09-364-425B-55 Sequence 55, App1
16	50.5	29.9	191	2	US-08-469-412A-4 Sequence 4, App1
17	50.5	29.9	191	3	US-09-021-715-4 Sequence 4, App1
18	50.5	29.9	209	4	US-08-311-731A-18 Sequence 18, App1
19	50.5	29.9	301	4	US-09-252-991A-18062 Sequence 18062, A
20	50	29.6	127	4	US-09-252-991A-26303 Sequence 26303, A
21	50	29.6	286	4	US-08-469-260A-275 Sequence 275, App
22	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
23	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
24	49.5	29.3	88	4	US-08-469-260A-275 Sequence 275, App
25	49.5	29.3	92	4	US-09-247-155-98 Sequence 98, App1
26	49.5	29.3	147	4	US-09-252-991A-23997 Sequence 23997, A
27	49.5	29.3	196	4	US-09-252-991A-31079 Sequence 31079, A

28	49.5	29.3	248	4	US-09-134-001C-5085 Sequence 5085, App
29	49.5	29.3	303	4	US-09-252-991A-17179 Sequence 17179, A
30	49	29.0	718	4	US-09-252-991A-25696 Sequence 25696, A
31	49	29.0	1498	4	US-09-252-991A-31234 Sequence 31234, A
32	48.5	28.7	251	4	US-09-252-991A-27585 Sequence 27585, A
33	48.5	28.7	534	4	US-09-252-991A-17265 Sequence 17265, A
34	48	28.4	170	4	US-09-148-545-204 Sequence 204, App
35	48	28.4	170	4	US-09-252-991A-26306 Sequence 26306, A
36	48	28.4	222	4	US-09-148-545-135 Sequence 135, App
37	48	28.4	229	4	US-09-252-991A-22195 Sequence 22195, A
38	48	28.4	246	4	US-09-252-991A-25102 Sequence 25102, A
39	48	28.4	290	4	US-09-489-039A-14080 Sequence 14080, A
40	48	28.4	337	4	US-09-252-991A-13602 Sequence 13602, A
41	48	28.4	439	4	US-09-489-039A-13336 Sequence 13336, A
42	48	28.4	495	4	US-09-489-039A-10528 Sequence 10528, A
43	48	28.4	718	4	US-09-657-960-3 Sequence 3, App1
44	48	28.4	1421	4	US-09-252-991A-17805 Sequence 17805, A
45	47.5	28.1	191	4	US-09-252-991A-25365 Sequence 25365, A

## ALIGNMENTS

```
RESULT 1
; US-09-252-991A-31463
; Sequence 31463, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,788
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31463
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31463

Query Match      32.5%; Score 55; DB 4; Length 109;
Best Local Similarity 42.1%; Pred. No. 2;
Matches      8; Conservative      2; Mismatches      9; Indels      0; Gaps      0;

QY      12 ASMCWALALLMLAVPGWS 30
Db      27 AAMCWSARSTWACSSAMS 45

RESULT 2
; US-09-252-991A-16906
; Sequence 16906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16906
; LENGTH: 263
; TYPE: PRT
```

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16906

Query Match 32.2%; Score 54.5; DB 4; Length 263;  
Best Local Similarity 46.4%; Pred. No. 5.9;  
Matches 13; Conservative 1; Mismatches 7; Indels 7; Gaps 2;

QY 3 RGAGAGRAGSMCWALALMLAVPGWS 30  
DB 23 RRGVARGCRARCMR---RW----PGWS 43

RESULT 3  
US-09-621-976-6090  
Sequence 6090, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 6090  
LENGTH: 74  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-621-976-6090

Query Match 31.4%; Score 53; DB 4; Length 74;  
Best Local Similarity 39.3%; Pred. No. 2.5;  
Matches 11; Conservative 3; Mismatches 10; Indels 4; Gaps 1;

QY 4 GAGAGRGA---SWCWALALMLAVP 27  
DB 13 GLGAPRGEGPGGWCWARSIFRCYP 40

RESULT 4  
US-09-220-528-104  
Sequence 104, Application US/09220528A  
Patent No. 6284540  
GENERAL INFORMATION:  
APPLICANT: Milbrandt, Jeffrey D.  
APPLICANT: Balon, Robert H.  
TITLE OF INVENTION: Artemin, A No. 6284540e1 Neurotrophic Factor  
FILE REFERENCE: 6029-7998  
CURRENT APPLICATION NUMBER: US/09/220,528A  
CURRENT FILING DATE: 1998-12-24  
EARLIER APPLICATION NUMBER: 09/218,698  
EARLIER FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 60/108,148  
EARLIER FILING DATE: 1998-11-12  
EARLIER APPLICATION NUMBER: 09/163,283  
EARLIER FILING DATE: 1998-09-29  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 104  
LENGTH: 215  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-220-528-104

Query Match 31.4%; Score 53; DB 3; Length 215;  
Best Local Similarity 35.7%; Pred. No. 7.6;  
Matches 10; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 2 RRGAGAGRAGSMCWALALMLAVPGW 29  
DB 73 QRRGAGAACARSMWCRCARSAWATPTSW 100

RESULT 5  
US-09-252-991A-32898  
Sequence 32898, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32898  
LENGTH: 228  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32898

Query Match 31.4%; Score 53; DB 4; Length 228;  
Best Local Similarity 36.7%; Pred. No. 8.1;  
Matches 11; Conservative 4; Mismatches 9; Indels 6; Gaps 2;

QY 2 RRGAGA-ARGRASMCWALALMLAVPGWS 30  
DB 25 QRRGPFHTRGCSATCW----WSAATPAMA 49

RESULT 6  
US-07-683-957B-2  
Sequence 2, Application US/07683957B  
Patent No. 5310880  
GENERAL INFORMATION:  
APPLICANT: Donahoe, Patricia K.  
APPLICANT: Ragin, Richard C.  
APPLICANT: MacLaughlin, David T.  
TITLE OF INVENTION: Purification of M 11erian Inhibiting  
SUBSTANCE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/683,957B  
FILING DATE: 19910412  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldstein, Jorge A.  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 0609,3060000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein





PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27810  
LENGTH: 370  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27810

Query Match 30.5%; Score 51.5; DB 4; Length 370;  
Best Local Similarity 38.5%; Pred. No. 22;  
Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

2y 7 AARGRAS--WCMALALMLAVPGM 29  
Db 112 AANSPASATWMTMTMTLAVM 137

RESULT 12  
US-09-431-705-22  
Sequence 22, Application US/09431705  
Patent No. 6585975  
GENERAL INFORMATION:  
APPLICANT: Kleantous, Harold  
APPLICANT: Londono-Arcila, Patricia  
APPLICANT: Freeman, Donna  
TITLE OF INVENTION: Use of salmonella vectors for  
FILE REFERENCE: 06132/060001  
CURRENT APPLICATION NUMBER: US/09/431,705  
CURRENT FILING DATE: 1999-11-01  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 85  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-431-705-22

Query Match 30.2%; Score 51; DB 4; Length 85;  
Best Local Similarity 52.9%; Pred. No. 5.4;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

2y 11 RASWCMALALMLAVP 27  
Db 16 QASWCMARRLVWMLASAP 32

RESULT 13  
US-09-205-258-274  
Sequence 274, Application US/09205258  
Patent No. 6525174  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 207 Human Secreted Proteins  
FILE REFERENCE: P200721  
CURRENT APPLICATION NUMBER: US/09/205,258  
CURRENT FILING DATE: 1998-12-04  
EARLIER APPLICATION NUMBER: PCT/US98/11422  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,885  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,375  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,881  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,880  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,896  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,020

EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,876  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,895  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,884  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,894  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,971  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,882  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,899  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,900  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,901  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,915  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,019  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,970  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,972  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,916  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,373  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,875  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/049,374  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,917  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,949  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,883  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,897  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,898  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,962  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,963  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,877  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/048,878  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/070,923  
EARLIER FILING DATE: 1997-12-18  
EARLIER APPLICATION NUMBER: 60/092,921  
EARLIER FILING DATE: 1998-07-15  
EARLIER APPLICATION NUMBER: 60/094,657  
EARLIER FILING DATE: 1998-07-30  
NUMBER OF SEQ ID NOS: 1227  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 274  
LENGTH: 88  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: SITE  
LOCATION: (53)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (88)  
OTHER INFORMATION: Xaa equals stop translation  
US-09-205-258-274

Query Match 30.2%; Score 51; DB 4; Length 88;  
Best Local Similarity 56.5%; Pred. No. 5.6;  
Matches 13; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

Y 1 MRRGA--GAAGRASMCALAL 21  
b 44 MAKGRGAKRGRRNGIAYTL 66

ESULT 14  
S-09-170-496D-290  
Sequence 290, Application US/09170496D  
Patent No. 6555339

## GENERAL INFORMATION:

APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-  
TITLE OF INVENTION: Receptors  
FILE REFERENCE: AREN-0040  
CURRENT APPLICATION NUMBER: US/09/170,496D  
CURRENT FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 294  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 290  
LENGTH: 1310  
TYPE: PRT  
ORGANISM: Homo sapiens  
S-09-170-496D-290

Query Match 30.2%; Score 51; DB 4; Length 1310;  
Best Local Similarity 52.9%; Pred. No. 95;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Y 11 RASMCWALALMLAVVP 27  
b 907 QASWCHARRLVWLHSAP 923

ESULT 15  
S-09-364-425B-55  
Sequence 55, Application US/09364425B  
Patent No. 6653086

## GENERAL INFORMATION:

APPLICANT: Behan, Dominic P.  
APPLICANT: Chalmers, Derek T.  
APPLICANT: Liaw, Chen W.  
APPLICANT: Lin, I-Lin  
APPLICANT: Lowitz, Kevin P.  
APPLICANT: Chen, Ruoping  
TITLE OF INVENTION: Endogenous, Constitutively Activated G Protein-Coupled Orphan Rece  
FILE REFERENCE: Aren0047  
CURRENT APPLICATION NUMBER: US/09/364,425B  
CURRENT FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: 60/094,879  
PRIOR FILING DATE: 1998-07-31  
PRIOR APPLICATION NUMBER: 60/106,300  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 60/110,906  
PRIOR FILING DATE: 1998-12-04  
PRIOR APPLICATION NUMBER: 60/121,851  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 55  
; LENGTH: 1310  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-364-425B-55

Query Match 30.2%; Score 51; DB 4; Length 1310;  
Best Local Similarity 52.9%; Pred. No. 95;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 11 RASMCWALALMLAVVP 27  
Db 907 QASWCHARRLVWLHSAP 923

Search completed: June 8, 2004, 14:09:16  
Job time: 2.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: June 8, 2004, 14:04:04 ; Search time 6.16667 Seconds

(without alignments)  
1368.675 Million cell updates/sec

File: US-10-010-050A-2\_COPY\_1\_30

Effect score: 169  
Sequence: 1 MRRGAGARGASWCWALALMLAVVPGMS 30

Coring table: BLASTSUM62  
Gapop 10.0, Gapext 0.5

Seached: 1155919 segs, 28138677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

Published Applications: AA.\*  
1: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubppa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubppa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubppa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubppa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubppa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubppa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169	100.0	346	US-09-122-383-2	Sequence 2, Appli
2	169	100.0	346	US-10-653-595-126	Sequence 126, App
3	169	100.0	346	US-10-010-050A-2	Sequence 2, Appli
4	169	100.0	346	US-09-397-945-126	Sequence 126, App
5	62	36.7	588	US-10-156-761-11538	Sequence 11538, A
6	60.5	35.8	192	US-09-811-284-167	Sequence 167, App
7	59.5	35.2	127	US-10-291-265-269	Sequence 269, App
8	58	34.3	188	US-09-989-920-240	Sequence 240, App
9	57	33.7	19608	US-10-084-846A-8	Sequence 8, Appli
10	56	33.1	214	US-10-080-170-198	Sequence 198, App
11	56	33.1	403	US-10-092-900A-300	Sequence 300, App
12	55.5	33.8	371	US-09-890-688-60	Sequence 60, App
13	53	31.4	28	US-09-242-450A-13	Sequence 13, Appli
14	53	31.4	65	US-09-242-450A-3	Sequence 3, Appli
15	53	31.4	117	US-10-094-749-2218	Sequence 2218, Ap

16	53	31.4	160	US-10-156-761-10933	Sequence 10933, A
17	53	31.4	215	US-09-220-920-104	Sequence 104, Appl
18	53	31.4	346	US-10-047-825-10	Sequence 10, Appl
19	53	31.4	368	US-10-425-114-55731	Sequence 55731, A
20	52.5	31.1	199	US-10-221-278-727	Sequence 727, App
21	52.5	31.1	199	US-10-291-172-727	Sequence 727, App
22	52.5	31.1	333	US-09-738-626-4771	Sequence 4771, Ap
23	52.5	31.1	360	US-10-282-122A-50396	Sequence 50396, A
24	52.5	31.1	464	US-10-424-598-207280	Sequence 207280, A
25	52	30.8	82	US-10-424-598-167251	Sequence 167251, A
26	52	30.8	560	US-09-863-776-69	Sequence 69, Appl
27	52	30.8	599	US-10-104-047-2306	Sequence 2306, Ap
28	52	30.8	653	US-09-863-776-28	Sequence 28, Appl
29	52	30.8	658	US-09-759-056-5	Sequence 5, Appli
30	52	30.8	9	US-09-901-812-5	Sequence 5, Appli
31	52	30.8	658	US-09-863-776-68	Sequence 68, Appl
32	52	30.8	662	US-09-863-776-30	Sequence 30, Appl
33	52	30.8	667	US-09-759-056-2	Sequence 2, Appli
34	52	30.8	667	US-09-901-812-2	Sequence 2, Appli
35	52	30.8	667	US-09-863-776-32	Sequence 32, Appl
36	52	30.8	667	US-09-863-776-67	Sequence 67, Appl
37	52	30.8	667	US-10-219-535-80	Sequence 80, Appl
38	52	30.8	667	US-10-232-230-80	Sequence 80, Appl
39	52	30.8	667	US-10-232-224-80	Sequence 80, Appl
40	52	30.8	667	US-10-227-884-80	Sequence 80, Appl
41	52	30.8	667	US-10-230-163-80	Sequence 80, Appl
42	52	30.8	667	US-10-230-338-80	Sequence 80, Appl
43	52	30.8	667	US-10-218-631-80	Sequence 80, Appl
44	52	30.8	667	US-10-230-414-80	Sequence 80, Appl
45	52	30.8	667	US-10-216-159A-80	Sequence 80, Appl

#### ALIGNMENTS

RESULT 1  
US-09-122-383-2  
; Sequence 2, Application US/09122383A  
; Patent No. US20020042093A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Gilbertson, Debra G.  
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
; FILE REFERENCE: 97-38  
; CURRENT APPLICATION NUMBER: US/09/122.383A  
; CURRENT FILING DATE: 1998-07-24  
; EARLIER APPLICATION NUMBER: 60/053,613  
; EARLIER FILING DATE: 1997-07-24  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-122-383-2

Query Match 100.0%; Score 169; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 3.8e-12;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRGAGARGASWCWALALMLAVVPGMS 30  
DB 1 MRRGAGARGASWCWALALMLAVVPGMS 30

RESULT 2  
US-10-653-595-126  
; Sequence 126, Application US/10653595  
; Publication No. US20040048304A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et. al.  
; TITLE OF INVENTION: 95 Human secreted proteins

```
FILE REFERENCE: P2027P1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIOR APPLICATION NUMBER: US 09/397945
PRIOR FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
S-10-653-595-126

Query Match          100.0%; Score 169; DB 12; Length 346;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

iy 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30
db 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30

RESULT 3
S-10-010-050A-2
Sequence 2, Application US/10010050A
Publication No. US20020173624A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38CI
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapien
S-10-010-050A-2
```

```
Query Match          100.0%; Score 169; DB 13; Length 346;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
iy 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30
db 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30
```

```
RESULT 4
US-09-397-945-126
Sequence 126, Application US/09397945
Publication No. US20030065139A1
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc. et al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1
CURRENT APPLICATION NUMBER: US/09/397,945
CURRENT FILING DATE: 1999-09-17
PRIOR APPLICATION NUMBER: PCT/US99/05804
PRIOR FILING DATE: 1999-03-18
PRIOR APPLICATION NUMBER: 60/078,566
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,576
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,573
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,574
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,579
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,314
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 126
LENGTH: 347
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (347)
OTHER INFORMATION: Xaa equals stop translation
US-09-397-945-126
```

```
Query Match          100.0%; Score 169; DB 12; Length 347;
Best Local Similarity 100.0%; Pred. No. 3,8e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
iy 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30
db 1 MRGAGAAAGRASWCWALALMLAVVPGWS 30
```

RESULT 5

```
S-10-156-761-11538
Sequence 11538, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHITAKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11538
LENGTH: 588
TYPE: PRT
ORGANISM: Streptomyces avermitilis
3-10-156-761-11538

Query Match          36.7%; Score 62; DB 14; Length 588;
Best Local Similarity 28.6%; Pred. No. 23;
Matches 14; Conservative 5; Mismatches 8; Indels 22; Gaps 1;

Y      3  RGAGAR-----GRASWCALALLMLAVPGW 29
      167  RGAGSRRTLYATKLTLLIGVGRYGDKSSWSWAGLIMLAPRARW 215

3-09-811-284-167
Sequence 167, Application US/09811284
Patent No. US20020058306A1
GENERAL INFORMATION:
APPLICANT: Vogel, Gabriel
TITLE OF INVENTION: No. US20020058306A1 G Protein-coupled Receptors
FILE REFERENCE: 00167U51
CURRENT APPLICATION NUMBER: US/09/811,284
CURRENT FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/189,783
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,907
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,918
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,960
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/189,917
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/192,945
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,916
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,923
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,933
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,830
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/192,234
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/192,155
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 60/192,935
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 258
SOFTWARE: PatentIn version 3.0

; SEQ ID NO 167
; LENGTH: 192
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-284-167

Query Match          35.8%; Score 60.5; DB 9; Length 192;
Best Local Similarity 39.3%; Pred. No. 13;
Matches 11; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

Y      4  GAGANRGRASWCW-----ALALLW 22
      35  GAAAVRGRPCWCPQCPALLVSIALLW 62

RESULT 7
US-10-291-265-269
; Sequence 269, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 269
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-269

Query Match          35.2%; Score 59.5; DB 15; Length 127;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 14; Conservative 3; Mismatches 9; Indels 9; Gaps 2;

Y      3  RGAGAR-----ASWCALALL--WLAVPG 28
      25  RGEETGRVWTWKLAQWLWGIALIGSTVALTTG 59

RESULT 8
US-09-989-920-240
; Sequence 240, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Heve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prote
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 240
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Homo sapien
```

US-09-989-920-240

Query Match 34.3%; Score 58; DB 9; Length 188;  
Best Local Similarity 38.7%; Pred. No. 26;  
Matches 12; Conservative 1; Mismatches 12; Indels 6; Gaps 1;

QY 4 GAGARGASWCM-----ALALMLAVPG 28  
DB 19 GGGGGAGAGSWTMMGSGGAGALMVAVGG 49

RESULT 9  
US-10-084-846A-8  
; Sequence 8, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEINBAUER, GABRIELE  
; APPLICANT: MUEHLMEIER, AGNES  
; APPLICANT: TREFFER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; PRIOR FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent In Ver. 3.2  
; SEQ ID NO 8  
; LENGTH: 19608  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-8

Query Match 33.7%; Score 57; DB 15; Length 19608;  
Best Local Similarity 42.9%; Pred. No. 1.8e+03;  
Matches 12; Conservative 3; Mismatches 11; Indels 2; Gaps 1;

Y 2 RRGAGARGASWCMALMLAVP 27  
DB 2876 RRAAVTPGARGRVMCGPVSGLVADP 2903

RESULT 10  
US-10-080-170-198  
; Sequence 198, Application US/10080170  
; Publication No. US20030129601A1  
; GENERAL INFORMATION:  
; APPLICANT: COLE, S.T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
; FILE REFERENCE: 03495.0218  
; CURRENT APPLICATION NUMBER: US/10/080,170  
; CURRENT FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270,123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 198  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-080-170-198

Query Match 33.1%; Score 56; DB 14; Length 214;  
Best Local Similarity 41.7%; Pred. No. 50;  
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 3;

QY 3 RGAGARGASWCMALMLAVP 26  
DB 175 RTSALARGITQMRMSIALMLFITV 198

RESULT 11  
US-10-092-900A-300  
; Sequence 300, Application US/10092900A  
; Publication No. US20040043382A1  
; GENERAL INFORMATION:  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Taupier Jr., Raymond J.  
; APPLICANT: Pena, Carol E.A.  
; APPLICANT: Li, Li  
; APPLICANT: Zethusen, Bryan D.  
; APPLICANT: Gusev, Vladimil Y.  
; APPLICANT: Ji, Weizhen  
; APPLICANT: Gorman, Linda  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Pakturajan, Weera  
; APPLICANT: Gangolli, Esna A.  
; APPLICANT: Vermet, Corine A.M.  
; APPLICANT: Guo, Xiaojia Sasha  
; APPLICANT: Tchermey, Velizar T.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Liu, Yi  
; APPLICANT: Anderson, David W.  
; APPLICANT: Spaderna, Steven K.  
; APPLICANT: Catterton, Elina  
; APPLICANT: Leite, Mario W.  
; APPLICANT: Zhong, Haihong  
; APPLICANT: Alsobrook, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: No. US20040043382A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-290C  
; CURRENT APPLICATION NUMBER: US/10/092,900A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: USSN 60/274,322  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/283,675  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: USSN 60/338,092  
; PRIOR FILING DATE: 2001-12-03  
; PRIOR APPLICATION NUMBER: USSN 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/274,191  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: USSN 60/325,681  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: USSN 60/304,354  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: USSN 60/279,995  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: USSN 60/294,899  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: USSN 60/287,424  
; PRIOR FILING DATE: 2001-04-30  
; Remaining Prior Application data removed - See File Wrapper or PAM.  
; NUMBER OF SEQ ID NOS: 768  
; SEQ ID NO 300  
; LENGTH: 403  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-092-900A-300

Query Match	33.1%;	Score 56;	DB 12;	Length 403;
Best Local Similarity	37.8%;	Pred. No. 86;		
Matches 14;	Conservative 1;	Mismatches 14;	Indels	

```

2y      2 RRAGAARGPASHWCNALALLMTAVPG-----WS 30
        |||  |||  |||  |||  |||  |||
2b      118 RRGISALRGAGLAWRLFLLAVTLPSIACILLYWS 154

```

RESULT 12  
NS-09-890-688-60  
Sequence 60, Application US/09890688  
Publication No. US2003014475A1  
GENERAL INFORMATION:

APPLICANT: Seishi KATO  
 APPLICANT: Chikashi EGUCHI  
 APPLICANT: Mihoro SAKKI  
 TITLE OF INVENTION: Human Proteins and cDNAs thereof  
 FILE REFERENCE: 2001-11024/WNC/00653  
 CURRENT APPLICATION NUMBER: US/09/590,688  
 CURRENT FILING DATE: 2001-08-06  
 PRIOR APPLICATION NUMBER: JP 11-346863  
 PRIOR FILING DATE: 1999-12-06  
 PRIOR APPLICATION NUMBER: JP 11-34684  
 PRIOR FILING DATE: 1999-12-06  
 PRIOR APPLICATION NUMBER: JP 2000-51062  
 PRIOR FILING DATE: 2000-02-08  
 PRIOR APPLICATION NUMBER: JP 2000-34091  
 PRIOR FILING DATE: 2000-02-10  
 PRIOR APPLICATION NUMBER: JP 2000-34090  
 PRIOR FILING DATE: 2000-02-10  
 PRIOR APPLICATION NUMBER: JP 2000-35829  
 PRIOR FILING DATE: 2000-02-14  
 PRIOR APPLICATION NUMBER: JP 2000-35899  
 PRIOR FILING DATE: 2000-02-14  
 PRIOR APPLICATION NUMBER: JP 2000-71161  
 PRIOR FILING DATE: 2000-03-14  
 PRIOR APPLICATION NUMBER: JP 2000-160851  
 PRIOR FILING DATE: 2000-05-30  
 NUMBER OF SEQ ID NOS: 160  
 SOFTWARE: PatentIn Ver. 2.1

Query Match	32.8%;	Score 55.5;	DB 10;	Length 371;
Best Local Similarity	55.6%;	Pred. No. 91;		
Matches 15; Conservative	0;	Mismatches 5;	Indels 7;	Gaps 2;

```

Y      3 RGAGCARGRASWCWALALMLAVP 29
      4 | | | | | | | | | |
      5 RAGGFARGLR---ALALAWL---PGW 24
      6 | | | | | | | | | |

```

RESULT 13  
 3-09-242-450A-13  
 Sequence 13, Application US/09242450A  
 Patent No. US20020155552A1  
 GENERAL INFORMATION:

APPLICANT: Bergmann, Johanna  
 TITLE OF INVENTION: "PRIONINS", HIGHLY SPECIFIC MARKERS FOR NONINVASIVE PRE-SYMPTOM  
 TITLE OF INVENTION: DETECTION OF TSE DISEASES AND TARGETS FOR THERAPEUTIC REAGENTS  
 TITLE OF INVENTION: CONTROL TSE DISEASES IN ANIMALS AND HUMANS  
 FILE REFERENCE: 930006-2001  
 CURRENT APPLICATION NUMBER: US/09/242,450A  
 CURRENT FILING DATE: 2000-02-16  
 NUMBER OF SEQ ID NOS: 13  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 13  
 LENGTH: 28

```

: TYPE: PRT
: ORGANISM: homo sapiens
:
: FEATURE:
: NAME/KEY: TRANSMEM
: LOCATION: (1)..(28)
:
: OTHER INFORMATION: membrane spanning helix of CUBs from amino acid 25 to amino acid
: OTHER INFORMATION: 52 in the complete sequenc
US-09-242-450A-13

```

Query Match	31.4%	Score 53;	DB 9;	Length 28;
Best Local Similarity	45.5%;	Pred. No. 20;		
Matches	10;	Conservative	1;	Mismatches 9; Indels

QY 10 GRASWCM--ALATLWLA VPGW 29  
| | | | : | | |  
Db 4 GAASWWLGAASWWLGAAPWW 25

```

RESULT 14
US-09-242-450A-3
; Sequence 3, Application US/09242450A
; Patent No. US2002015552A1
; GENERAL INFORMATION:
; APPLICANT: Bergmann, Johanna
; TITLE OF INVENTION: "PRIONINS", HIGHLY SPECIFIC MARKERS FOR NONINVASIVE PRE-SYMPTOM
; TITLE OF INVENTION: DETECTION OF TSE DISEASES AND TARGETS FOR THERAPEUTIC REAGENTS 1
; TITLE OF INVENTION: CONTROL TSE DISEASES IN ANIMALS AND HUMANS
; FILE REFERENCE: 630006-2001
; CURRENT APPLICATION NUMBER: US/09/242,450A
; CURRENT FILING DATE: 2000-02-16
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 65
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-242-450A-3

```

Query Match	31.4%;	Score 53;	DB 9;	Length 65;
Best Local Similarity	45.5%;	Pred. No. 41;		
Matches 10; Conservative	1;	Mismatches 9;	Indels 2;	Gaps 1;

QY 10 GRASWCW--ALALLWLAVVPGW 29  
| | | | : | | |  
Db 28 GAASWWLGAASWWLGAAPWW 49

```

US/10-094-749-2218
; Sequence 2218, Application US/10094749
; Publication No. US20030219741A1
GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTAKO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOMOYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
FILE REFERENCE: 084353/0160
CURRENT APPLICATION NUMBER: US/10/094,749

```



```

; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2218
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2218

```

```

Query Match      31.4%; Score 53; DB 15; Length 117;
Best Local Similarity 36.6%; Pred. No. 67;
Matches 15; Conservative 2; Mismatches 12; Indels 12; Gaps 2;

2y      2  RRGAGA---ARGRAS-----WCWALALMLLAVPGWS 30
          ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
          54  RRGAGATGARGRGHGHSSLIAGPWRMLLSFSEFEKVVNWS 94

```

Search completed: June 8, 2004, 14:11:57  
 Job time : 6.16667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: June 8, 2004, 13:59:38 ; Search time 2.16667 Seconds

(without alignments)  
1331.883 Million cell updates/sec

title: US-10-010-050A-2\_COPY\_1\_30

effect score: 169

equence: 1 MRRGAGARGRASWCALALMLAVPGWS 30

oring table: BLOSUM62

earch: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

#### SUMMARIES

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	66	39.1	606	2 C84390	sulfate transport
2	59	34.9	412	2 D95941	conserved hypotet
3	56	33.1	214	2 T44701	probable integral
4	55	32.5	232	2 F82729	conserved hypotet
5	53.5	31.7	159	2 A48428	homeotic protein H
6	53.5	31.7	284	2 A82729	phosphatidate cycl
7	53	31.4	398	2 S11800	pululanase secret
8	53	31.4	428	2 A12860	conserved hypotet
9	53	31.4	428	2 H97637	hypothetical prote
10	52.5	31.1	220	2 AG3627	maltoase transport
11	52.5	31.1	575	1 WFBOM	mullerian inhibiti
12	52	30.8	641	2 A45054	probable intercell
13	51.5	30.5	686	2 S30075	ferric reductase (
14	51.5	30.5	1146	2 A38587	collagen, cornea-s
15	51	30.2	410	2 E35984	conserved hypotet
16	50.5	29.9	210	2 E70671	hypothetical prote
17	50	29.6	106	2 A29760	hypothetical prote
18	50	29.6	163	2 S73039	hypothetical prote
19	50	29.6	201	2 D81944	probable integral
20	50	29.6	219	2 S70311	hypothetical prote
21	50	29.6	221	2 G83964	hypothetical prote
22	50	29.6	249	2 B84147	ABC transporter (p
23	50	29.6	250	2 AD1197	ABC transporter tr
24	50	29.6	363	2 T34631	probable integral
25	50	29.6	417	2 AF3448	hypothetical membr
26	50	29.6	514	2 S17958	cytochrome oxidase
27	50	29.6	514	2 F90770	probable third cyt
28	50	29.6	514	2 B85633	probable third cyt
29	50	29.6	529	2 AF3059	cytochrome d oxida

30	50	29.6	529	2 B98227	cytochrome d oxida
31	49.5	29.3	310	2 T16233	hypothetical prote
32	49.5	29.3	519	2 S75570	apolipoprotein N-a
33	49	29.0	118	2 S52855	hypothetical prote
34	49	29.0	139	2 A71123	hypothetical prote
35	49	29.0	188	2 B82183	ankb protein VC158
36	49	29.0	250	2 AB1555	ABC transporter tr
37	49	29.0	317	2 T35010	probable integral
38	49	29.0	755	2 B75346	probable competenc
39	49	29.0	1221	2 E83327	conserved hypotet
40	48.5	28.7	111	2 T14306	glycine-rich prote
41	48.5	28.7	125	2 C83138	hypothetical prote
42	48.5	28.7	210	2 S16297	T-cell receptor ga
43	48.5	28.7	276	2 B82243	hypothetical prote
44	48.5	28.7	435	2 A12082	hypothetical prote
45	48.5	28.7	688	2 B97152	probable membrane

#### ALIGNMENTS

##### RESULT 1

C84390 sulfate transport system permease protein [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84390

R:Ng, M.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebdhardt, H.; Lowe, T.M.; L

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: C84390

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-606 <STO>

A:Cross-references: GB:AB004437; NID:g10581803; PID:AA620487.1; GSPDB:GN00138

C:Gene: cyst1

C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001

C:Accession: D95941

R:Finan, T.W.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.U.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,663-kb P<sub>2</sub> megaplasmid from the N<sub>2</sub>-fixing endo

A:Reference number: A35842; MUID:21396508; PMID:11481431

A:Accession: D95941

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-412 <KOR>

A:Cross-references: GB:AL591985; PID:CA949196.1; PID:g15140681; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeLaure,

hebaul, P.; Vandenbol, M.; Vorholter, F.U.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: SMD21292  
 A:Genome: plasmid

Query Match 34.9%; Score 59; DB 2; Length 412;  
 Best Local Similarity 37.0%; Pred. No. 3.1;  
 Matches 10; Conservative 6; Mismatches 5; Indels 6; Gaps 1;

QY 7 AARGRASMCMW-----LALLMLAVP 27  
 DB 276 ANKGRPTWTLATVLFALIMLSTVP 302

## RESULT 3

144701  
 Probable integral membrane protein [imported] - *Mycobacterium leprae*

C:Species: *Mycobacterium leprae*  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 28-Jul-2000  
 C:Accession: 144701  
 R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, May 1998  
 A:Reference number: 222830  
 A:Accession: 144701

A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-214 <PAR>  
 A:Cross-references: EMBL:AL023635; PIDN:CAA19188.1  
 A:Experimental source: cosmid B1243  
 A:Genetics:  
 A>Note: MLCB1243.07  
 A:Superfamily: *Mycobacterium tuberculosis* hypothetical protein Ry2968c

Query Match 33.1%; Score 56; DB 2; Length 214;  
 Best Local Similarity 41.7%; Pred. No. 4.4;  
 Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 3 RGAGARGRSMCMWALALLMLAVV 26  
 DB 175 RTSALARGIYQWRWSTATLMTITV 138

## RESULT 4

82729

Conserved hypothetical protein XF1054 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
 C:Accession: F82729

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequence  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: F82729

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-232 <SIM>  
 A:Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83864.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 rtiões, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.F.A.; Carraro, D.M.; Carreir, H  
 e-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.U.S.  
 Submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme  
 D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre  
 hado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaask  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 ; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1054  
 C:Superfamily: *Streptomyces coelicolor* probable integral membrane protein SC6G10.12

Query Match 32.5%; Score 55; DB 2; Length 232;  
 Best Local Similarity 44.0%; Pred. No. 6.3;  
 Matches 11; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 8 ARGRASCMALALLMLAVPG--WS 30  
 DB 13 SRNPTWMLAVPLFFAVLIFLWS 37

## RESULT 5

A48428  
 Homeotic protein Hox 1.7 (clone MH-1) - guinea pig (fragment)

C:Species: *Cavia porcellus* (guinea pig)  
 C:Date: 19-Nov-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Oct-1997  
 C:Accession: A48428  
 R:Rubin, M.R.; Nguyen-Huu, M.C.  
 DNA Seq. 1, 115-124, 1990

A:Title: Alternatively spliced Hox-1.7 transcripts encode different protein products.  
 A:Reference number: A48428; MUID:92190539; PMID:1983703  
 A:Accession: A48428  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-159 <RUB>  
 A>Note: sequence extracted from NCBI backbone (NCBI:89505, NCBI:89507)  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:94-150/Domain: homeobox homology <HDX>

Query Match 31.7%; Score 53.5; DB 2; Length 159;  
 Best Local Similarity 44.4%; Pred. No. 7.2;  
 Matches 12; Conservative 3; Mismatches 9; Indels 3; Gaps 1;

QY 6 GAARGRSMCMW---ALALLMLAVPGW 29  
 DB 10 GSSRSRSMCMNKLQALFAGLADPAW 36

## RESULT 6

A82729

phosphatidate cytidyltransferase XF1049 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: A82729

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequence  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: A82729

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-284 <SIM>  
 A:Cross-references: GB:AE003942; GB:AE003849; NID:99105990; PIDN:AAF83859.1; GSPDB:GN001;  
 A:Experimental source: strain 9a5c  
 R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 rtiões, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.F.A.; Carraro, D.M.; Carreir, H  
 e-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.U.S.  
 Submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme  
 D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaask  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z



A:Reference number: A90879; MUID:86218082; PMID:3754790  
A:Accession: A01398  
A:Molecule type: DNA  
A:Residues: 1-14 <CAI>  
A:Experimental source: newborn calf testis, clones cblm15 and ps21  
A:Accession: B01398  
A:Molecule type: mRNA  
A:Residues: 15-575 <CA2>  
C:Comment: This glycoprotein, produced by the Sertoli cells of the testis, causes regression of Mullerian duct origin. Other roles for this protein in gonadal differentiation, testis duct regression and in the adult ovary.  
C:Comment: This protein is homologous to the beta transforming growth factor, inhibin A, these sequences. All of these proteins are biologically active as disulfide-linked dimers.  
C:Comment: Although it does not compete with EGF for receptor binding sites, MIS can inhibit C-Superfamily: inhibin  
C:Keywords: cytotrophin; glycoprotein; gonadal differentiation; testis  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-24/Domain: propeptide #status predicted <PRO>  
F:25-575/Product: mullerian inhibiting factor #status predicted <MAT>  
F:18,344/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match  
Best Local Similarity 31.1%; Score 52.5; DB 1; Length 575;  
Matches 11; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

4 GAGAGRGRASWCMALMLAVVP 28  
41 GSGALIFQQAQMDPUSLWL--PG 62

RESULT 12  
A45054  
Probable intercellular signal transducer or transmitter Fz-1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Mar-2001  
C:Accession: A45054  
R:Chan, S.D.; Karpf, D.B.; Fowlkes, M.R.; Hooks, M.; Bradley, M.S.; Vuong, V.; Bambino, J. Biol. Chem. 267, 25202-25207, 1992  
A:Title: Two homologs of the Drosophila polarity gene frizzled (Fz) are widely expressed  
A:Reference number: A45054; MUID:93094228; PMID:1334084  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-641 <CHAS>  
A:Experimental source: UMR 106 osteosarcoma cell line  
A>Note: Sequence extracted from NCBI Backbone (NCBIP:120154)  
C:Superfamily: fruit fly frizzled protein

Query Match  
Best Local Similarity 30.8%; Score 52; DB 2; Length 641;  
Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

5 AGAAGRA-SWCMA---LALMLAVP 27  
38 AGHRPRRAHRCWARGLLMLLEAP 64

RESULT 13  
S30075  
Fertic reductase (EC 1.6.99.-) FRE1 - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein L8167.2; protein YLR214W  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-May-1993 #sequence\_revision 28-May-1993 #text\_change 17-Mar-2000  
C:Accession: S30075; S48565  
R:Dancs, A.; Roman, D.G.; Anderson, G.T.; Hinebuch, A.G.; Klausner, R.D.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3869-3873, 1992  
A:Title: Fertic reductase of Saccharomyces cerevisiae: Molecular characterization, role  
A:Reference number: S30075; MUID:92237270; PMID:1570306  
A:Accession: S30075  
A:Molecule type: DNA  
A:Residues: 1-686 <DAN>  
A:Cross-references: EMBL:M86908; NID:G171520; PIDN:AAA34608.1; PID:G171521  
R:Pauley, A.

submitted to the EMBL Data Library, September 1994  
A:Description: The sequence of S. cerevisiae cosmid 8167.  
A:Reference number: S48545  
A:Accession: S48565  
A:Molecule type: DNA  
A:Residues: 1-686 <PAU>  
A:Cross-references: EMBL:U14913; NID:G544497; PIDN:AAB67424.1; PID:G544499; MIPS:YLR214W  
C:Genetic:  
A:Gene: SGD:PRE1  
A:Cross-references: SGD:S0004204; MIPS:YLR214W  
A:Map position: 12R  
C:Superfamily: ferric reductase FRE2  
C:Keywords: oxidoreductase; transmembrane protein  
F:6-22/Domain: transmembrane #status predicted <TM1>  
F:152-168/Domain: transmembrane #status predicted <TM2>  
F:216-232/Domain: transmembrane #status predicted <TM3>  
F:258-274/Domain: transmembrane #status predicted <TM4>  
F:298-314/Domain: transmembrane #status predicted <TM5>  
F:329-345/Domain: transmembrane #status predicted <TM6>  
F:359-375/Domain: transmembrane #status predicted <TM7>  
F:531-547/Domain: transmembrane #status predicted <TM8>

Query Match  
Best Local Similarity 30.5%; Score 51.5; DB 2; Length 686;  
Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

11 RASNC-WALMLMLAVP 26  
144 RSQWCAMGLVFWVAVL 160

Query Match  
Best Local Similarity 30.5%; Score 51.5; DB 2; Length 1146;  
Matches 15; Conservative 2; Mismatches 6; Indels 13; Gaps 3;

4 GAGAGRA-SWC-----WALML--MLAVP 26  
29 GAGATGAPSWCPGSCGSMWKMGLLMLMLL 64

;Residues: 1-410 <KUR>  
 ;Cross-references: GB:AL591985; PTDN:CAC49541.1; PID:G15141028; GSPDB:GN00167  
 ;Experimental source: strain 1021, megaplasmid pSymB  
 ;Gailbert, F.; Fhan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 J.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 ;Science 293, 668-672, 2001  
 ;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 J.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 ;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 ;Reference number: A96039; MUID:21368234; PMID:11474104  
 ;Contents: annotation  
 ;Genetics:  
 ;Gene: Smb20848  
 ;Genome: plasmid

Query Match 30.2%; Score 51; DB 2; Length 410;  
 Best Local Similarity 29.6%; Pred. No. 33;  
 Matches 8; Conservative 6; Mismatches 7; Indels 6; Gaps 1;  
 Y 7 AARGASWCWA-----LALLMLAVP 27  
 b 276 ARKGRPTWTWLVTTLEFILIMLSTAP 302

Search completed: June 8, 2004, 14:08:09  
 Job time: 3.16667 secs



DR MIM; 256731; -  
 DR GO; GO:0016021; C: integral to membrane; TAS.  
 DR GO; GO:0008151; P: cell growth and/or maintenance; TAS.  
 KW Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;  
 KW disease mutation; Polymorphism; Epilepsy.  
 FT TRANSMEM 75 91  
 FT CARBOHYD 179 179  
 FT CARBOHYD 192 192  
 FT CARBOHYD 227 227  
 FT CARBOHYD 252 252  
 FT CARBOHYD 304 304  
 FT CARBOHYD 320 320  
 FT CARBOHYD 330 330  
 FT CARBOHYD 401 401  
 FT VARIANT 279  
 FT VARIANT 368 368  
 FT SEQUENCE 407 AA, 4639 MW, 449702D1DC9BEE4 CRC64;  
 Query Match 100.0%; Score 169; DB 1; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 1,4e-14;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2Y 1 MRGAGARGASWCWALALMLAVVPGWS 30  
 62 MRGAGARGASWCWALALMLAVVPGWS 91  
 RESULT 2  
 CRCB\_PALSO STANDARD; PRT; 126 AA.  
 NC Q8XZR2;  
 JT 28-FEB-2003 (Rel. 41, Created)  
 JT 28-FEB-2003 (Rel. 41, Last sequence update)  
 JT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein crcb homolog.  
 IN CRCB OR RSC133 OR RS02855.  
 NC Ralstonia solanacearum (Pseudomonas solanacearum).  
 NC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 NC Burkholderiaceae; Ralstonia.  
 NC NCBI\_TaxId=305;  
 IN [1]  
 DE SEQUENCE FROM N.A.  
 NC STRAIN=GMT1000;  
 NC MEDLINE=21681879; PubMed=11823852;  
 UA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 UA Aclat M., Billault A., Brotier P., Camus J.C., Catolico L.,  
 UA Gaspin C., Lavie M., Molan A., Robert C., Saurin W., Schlex T.,  
 UA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 UA Weissenbach J., Boucher C.A.;  
 JT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 JT Nature 415:497-502(2002).  
 NC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 NC -1- SIMILARITY: Belongs to the crcb family.  
 NC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 NC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 NC the European Bioinformatics Institute. There are no restrictions on its  
 NC use by non-profit institutions as long as its content is in no way  
 NC modified and this statement is not removed. Usage by and for commercial  
 NC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 NC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 NC EMBL; AL646064; CAD15035.1; -  
 NC HAMAP; MF\_00454; -; 1.  
 NC InterPro; IPR003691; Camphor\_CrcB.  
 NC Pfam; PF02537; CRCB; 1.  
 NC TIGRPFAM; TIGR00494; crcb; 1.  
 KW Transmembrane; Complete proteome.  
 T TRANSMEM 5 24  
 T TRANSMEM 34 56  
 T POTENTIAL.  
 T POTENTIAL.

FT TRANSMEM 68 90  
 FT TRANSMEM 100 122  
 FT SEQUENCE 126 AA, 13101 MW, C2443FBA5C81CB3 CRC64;  
 Query Match 33.4%; Score 56.5; DB 1; Length 126;  
 Best Local Similarity 52.0%; Pred. No. 1.3;  
 Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;  
 QY 4 GAGAGARGASWCWALALMLAVV 27  
 DB 10 GVGAALG-AWLNRMAFVLMNAIN 32  
 RESULT 3  
 HXA9\_CAVPO STANDARD; PRT; 162 AA.  
 ID HXA9\_CAVPO  
 AC P51783;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-A9 (Hox-1.7) (Fragment).  
 GN HOXA9 OR HOX-1.7.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae; Cavia.  
 NC NCBI\_TaxId=10141;  
 RN [1]  
 RE SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA MEDLINE=92190539; PubMed=1983703;  
 RX Rubin M.R., Nguyen-Huu M.C.;  
 RT "Alternatively spliced Hox-1.7 transcripts encode different protein  
 RT products.";  
 RL DNA Seq. 1:115-124(1990).  
 CC -1- FUNCTION: Sequence-specific transcription factor which is part of  
 CC a developmental regulatory system that provides cells with  
 CC specific positional identities on the anterior-posterior axis.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the Abd-B homeobox family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X13536; CAA11887.1; -  
 CC HSSP; P02833; 9ANT.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR006711; Hox9 act.  
 DR InterPro; IPR000047; HTH\_lambdarepressr.  
 DR Pfam; PF00046; homeobox; 1.  
 DR Pfam; PF04617; Hox9 act; 1.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR PRINTS; PR00031; HTHREPRESSR.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT NON\_TER 1  
 FT DNA\_BIND 96 155  
 FT SEQUENCE 162 AA, 18661 MW, B84FA507C67CF91 CRC64;  
 Query Match 31.7%; Score 53.5; DB 1; Length 162;  
 Best Local Similarity 44.4%; Pred. No. 3.9;  
 Matches 12; Conservative 3; Mismatches 9; Indels 3; Gaps 1;  
 QY 6 GAGARGASWCW--ALALMLAVVPGW 29  
 ALALMLAVVPGW 29





```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Apple Hill;
RX      MEDLINE=93012921; PubMed=1398053;
RA      Carulli J.P., Hartl D.L.;
RT      "Variable rates of evolution among Drosophila opsin genes.";
RL      Genetics 132:193-204(1992).
CC      -!- FUNCTION: Visual pigments are the light-absorbing molecules that
CC      mediate vision. They consist of an apoprotein, opsin, covalently
CC      linked to cis-retinal.
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -!- PFM: Some or all of the carboxyl-terminal Ser or Thr residues may
CC      be phosphorylated.
CC      -!- MISCELLANEOUS: Each Drosophila eye is composed of 800 facets or
CC      ommatidia. Each ommatidium contains 8 photoreceptor cells (R1-R8),
CC      the R1 to R6 cells are outer cells, while R7 and R8 are inner
CC      cells.
CC      -!- MISCELLANEOUS: Opsin Rh1 has an absorption maximum at 480 nm.
CC      -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC      Opsin subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X65877; CAA46708.1; -.
CC      PIR: S40691; S40691.
CC      HSSP: P02699; 1EDV.
CC      PDBase: FBgn0012733; Dpse\minaE.
CC      InterPro: IPR000276; GPCR_Rhodopsn.
CC      InterPro: IPR001760; Opsin.
CC      Pfam: PF00001; 7tm.1; 1.
CC      PRINTS: PR00237; GPCR_Rhodopsn.
CC      PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
CC      PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
CC      PROSITE: PS00238; OPSIN; 1.
CC      W: Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
CC      Glycoprotein; G-protein coupled receptor; Vision.
CC      W: Glycoprotein; G-protein coupled receptor; Vision.
CC      T: TRANSMEM 1 49 EXTRACELLULAR.
CC      T: DOMAIN 50 74 1 (POTENTIAL).
CC      T: TRANSMEM 75 86 CYTOPLASMIC.
CC      T: TRANSMEM 87 112 2 (POTENTIAL).
CC      T: DOMAIN 113 126 EXTRACELLULAR.
CC      T: TRANSMEM 127 146 3 (POTENTIAL).
CC      T: DOMAIN 147 165 CYTOPLASMIC.
CC      T: TRANSMEM 166 190 4 (POTENTIAL).
CC      T: DOMAIN 191 214 5 (POTENTIAL).
CC      T: TRANSMEM 215 242 6 (POTENTIAL).
CC      T: TRANSMEM 243 277 7 (POTENTIAL).
CC      T: DOMAIN 278 301 8 (POTENTIAL).
CC      T: TRANSMEM 302 308 EXTRACELLULAR.
CC      T: TRANSMEM 309 333 7 (POTENTIAL).
CC      T: DOMAIN 334 374 CYTOPLASMIC.
CC      T: CARBOHYD 374 374 CYTOPLASMIC.
CC      T: DISULFID 20 20 N-LINKED (GLCNAC. . .) (PROBABLE).
CC      T: BINDING 123 201 RETINAL.
CC      T: BINDING 320 320 RETINAL CHROMOPHORE.
CC      Q: SEQUENCE 374 AA, 41604 MW, 8228082D97E0FA34 CRC64;
CC
CC      Query Match 30.8%; Score 52; DB 1; Length 374;
CC      Best Local Similarity 32.0%; Pred. No. 13;
CC      Matches 8; Conservative 7; Mismatches 8; Indels 2; Gaps 1;
CC
CC      Y 8 ARGRASWCALALMLAVVP--GMS 30
CC      | : : : : : |
CC      b 166 ALGKIAYITMFSTIMCCLAPVPGWS 190

```

ID	FDZ1 RAT	STANDARD;	PRT;	641 AA.
AC	008463;			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Frizzled 1 precursor (Frizzled-1) (Fz-1) (Fz1).			
GN	FDZ1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI_FaxID	101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Sprague-Dawley; TISSUE=Osteosarcoma;			
RX	MEDLINE=93094228; PubMed=1334084;			
RA	Chan S.D.H., Karpf D.B., Fowlkes M.E., Hooks M., Bradley M.S., Vuong V., Bambino T., Liu M.Y.C., Arnold C.D., Stewler G.J., Nissensohn R.A.;			
RT	"Two homologs of the Drosophila polarity gene frizzled (fz) are widely expressed in mammalian tissues.";			
RL	J. Biol. Chem. 267:25202-25207(1992).			
RN	[2]			
RP	COUPLING TO BETA-CATENIN PATHWAY.			
RX	MEDLINE=9924245; PubMed=10395542;			
RA	Sheldahl L.C., Park M., Malbon C.C., Moon R.T.;			
RT	"Protein kinase C is differentially stimulated by Wnt and Frizzled homologs in a G-protein-dependent manner.";			
RL	Curr. Biol. 9:695-698(1999).			
CC	-1- FUNCTION: Receptor for Wnt proteins. Most of frizzled receptors are coupled to the beta-catenin canonical signaling pathway, which leads to the activation of dishevelled proteins, inhibition of GSK-3 kinase, nuclear accumulation of beta-catenin and activation of Wnt target genes. A second signaling pathway involving PKC and calcium fluxes has been seen for some family members, but it is not yet clear if it represents a distinct pathway or if it can be integrated in the canonical pathway, as PKC seems to be required for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem to involve interactions with G-proteins. May be involved in transduction and intercellular transmission of polarity information during tissue morphogenesis and/or in differentiated tissues. Activation by Wnt8 induces expression of beta-catenin target genes.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- TISSUE SPECIFICITY: Widely expressed. Most abundant in kidney, liver, uterus, ovary and heart. Lower levels seen in brain and intestine. Extremely low in calvaria, mammary glands and testis.			
CC	-1- DEVELOPMENTAL STAGE: Expressed predominantly in neonatal tissues, at lower levels in adult.			
CC	-1- DOMAIN: Lys-Thr-X-X-X-Tip motif is involved in the activation of the Wnt/beta-catenin signaling pathway (By similarity).			
CC	-1- DOMAIN: The Fz domain is involved in binding with Wnt ligands (By similarity).			
CC	-1- SIMILARITY: Belongs to the Fz/Smo G-protein coupled receptor family.			
CC	-1- SIMILARITY: Contains 1 frizzled (Fz) domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL, L02529; AAA1173.1; -			
DR	InterPro: IPR000539; Frizzled.			
DR	InterPro: IPR000024; Fz domain.			
DR	InterPro: IPR000833; GPCR secretin.			
DR	Pfam: PF01534; Frizzled, I.			
DR	Pfam: PF01392; Fz, 1			
DR	PRINTS, PR00489; FRIZZLED.			
DR	SMART, SW00063; FRI, 1.			
DR	PROSITE, PS50038; FZ, 1.			

DR PROSITE; P550261; G PROTEIN RECEPTOR F2\_4; 1.  
 KM Multigene family; G-protein coupled receptor; Transmembrane;  
 KM Developmental Protein; Wnt signaling pathway; Glycoprotein; Signal.  
 FT SIGNAL 1 68  
 FT CHAIN 69 641  
 FT DOMAIN 69 316  
 FT TRANSMEM 317 337  
 FT DOMAIN 338 348  
 FT TRANSMEM 349 369  
 FT DOMAIN 370 396  
 FT TRANSMEM 397 417  
 FT DOMAIN 418 439  
 FT TRANSMEM 440 460  
 FT DOMAIN 461 483  
 FT TRANSMEM 484 504  
 FT DOMAIN 505 530  
 FT TRANSMEM 531 551  
 FT DOMAIN 552 595  
 FT TRANSMEM 596 616  
 FT DOMAIN 617 641  
 FT DOMAIN 106 224  
 FT SITE 85 90  
 FT SITE 619 624  
 FT CARBOHYD 125 125  
 FT CARBOHYD 225 225  
 FT SEQUENCE 641 AA; 71027 MW; D82E2C13E81B8B6 CRC64;  
 Query Match 30.5%; Score 52; DB 1; Length 641;  
 Best Local Similarity 55.6%; Pred. No. 20;  
 Matches 15; Conservative 0; Mismatches 8; Indels 4; Gaps 2;  
 Y 5 AGAARGRA-SWCNA---LALLMLAVVP 27  
 b 38 AGHRRPRARRHRCVARGLLMLLEAP 64  
 RESULT 8  
 REL\_YEAST STANDARD; PRT; 686 AA.  
 D FREI\_YEAST P32791;  
 C 01-OCT-1993 (Rel. 27, Created)  
 T 01-OCT-1993 (Rel. 27, Last sequence update)  
 T 10-OCT-2003 (Rel. 42, Last annotation update)  
 E Ferric reductase transmembrane component 1 precursor (BC 1.16.1.7)  
 N (Ferric-chelate reductase 1).  
 FREL OR YLR214W OR L8167.2.  
 S Saccharomyces cerevisiae (Baker's yeast).  
 C Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 C Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 K NCBI\_TaxID=4932;  
 N (1)  
 P SEQUENCE FROM N.A.  
 K STRAIN=F113;  
 K MEDLINE=92237270; PubMed=1570306;  
 K Dancis A., Roman D.G., Anderson G.J., Hinnebusch A.G., Klausner R.D.;  
 "Ferric reductase of *Saccharomyces cerevisiae*: molecular  
 characterization, role in iron uptake, and transcriptional control by  
 iron.";  
 J Proc. Natl. Acad. Sci. U.S.A. 89:3869-3873(1992).  
 [2]  
 2 SEQUENCE FROM N.A.  
 3 STRAIN=S288C / AB972;  
 4 MEDLINE=97313267; PubMed=9169871;  
 5 Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,  
 6 Benes V., Bruckner M., Delius H., Dubois E., Duesterhoeft A.,  
 7 Entian K.-D., Floeth M., Goffeau A., Hebling U., Henmann K.,  
 8 Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koelter P.,  
 9 Louis E.-J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,  
 10 Mueller-Auer S., Netuschil U., Obermaier B., Piravandi E., Pohl T.M.,  
 11 Portetelle D., Purnelle B., Reumann S., Rieger M., Rinke M., Rose M.,  
 12 Schaefer M., Scherens B., Scholler P., Schwaiger C., Schwarz S.,  
 13 Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,

RA Viereendeels F., Voet M., Volckaert G., Voss H., Wardutt R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Han J., Hohnsels J.D.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII.";  
 RT Nature 387:87-90(1997).  
 RN (3)  
 RN CHARACTERIZATION.  
 RP MEDLINE=93057491; PubMed=1431884;  
 RA Anderson G.J., Lesuisse E., Dancis A., Roman D.G., Labbe P.,  
 RA Klausner R.D.;  
 RT "Ferric iron reduction and iron assimilation in *Saccharomyces*  
 RT *cerevisiae*.";  
 RL J. Inorg. Biochem. 47:249-255(1992).  
 CC -1- FUNCTION: Reductase activity that acts on ferric iron chelates  
 CC external to the cell. Plays a role in iron uptake. May  
 CC participate in the transport of electrons from cytoplasm to an  
 CC extracellular substrate (ferric ion) via FAD and heme  
 CC intermediates. May also participate in Cu(II) reduction and Cu(I)  
 CC uptake.  
 CC -1- CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).  
 CC -1- COFACTOR: FAD (Probable).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- INDUCTION: BY IRON DEPRIVATION. REPRESSED BY IRON UPTAKE.  
 CC -1- SIMILARITY: Belongs to the FRE / CYBB family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M66908; AAA34608.1; -;  
 CC EMBL; U14913; AAB67424.1; -;  
 CC PIR; S30075; S30075.  
 CC Germonline; 142276; -;  
 CC DR SGD; S0004204; FRE1.  
 CC DR GO; GO:0005886; Cytoplasm membrane, IDA.  
 CC DR GO; GO:000293; Ferric-chelate reductase activity, IDA.  
 CC DR GO; GO:0015677; P-copper ion import, IDA.  
 CC DR GO; GO:0006826; P-iron ion transport, IDA.  
 CC DR InterPro; IPR002916; Ferric\_reduct.  
 CC DR Pfam; PF01794; Ferric\_reduct.  
 CC KM Oxidoreductase, Electron transport; Transmembrane; Iron transport;  
 CC FAD; NAD; Copper; Glycoprotein; Signal; Multigene family.  
 CC FT SIGNAL 1 22  
 CC CHAIN 23 686  
 CC FT NP BIND 462 468  
 CC FT NP BIND 532 540  
 CC FT TRANSMEM 147 169  
 CC FT TRANSMEM 216 236  
 CC FT TRANSMEM 258 277  
 CC FT TRANSMEM 296 316  
 CC FT TRANSMEM 329 348  
 CC FT TRANSMEM 353 397  
 CC FT TRANSMEM 529 550  
 CC FT CARBOHYD 69 69  
 CC FT CARBOHYD 100 100  
 CC FT CARBOHYD 124 124  
 CC FT CARBOHYD 209 209  
 CC FT CARBOHYD 503 503  
 CC FT CARBOHYD 646 646  
 CC SEQUENCE 686 AA; 78853 MW; 756BB3B9A95D6A3 CRC64;  
 Query Match 30.5%; Score 51.5; DB 1; Length 686;  
 Best Local Similarity 47.1%; Pred. No. 25;  
 Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;  
 QY 11 RASWC-WALLMLAVVP 26  
 Db 144 RSCWCAWGLVFFWAVL 160

```

RESULT 9
RT SH_RAT STANDARD; PRT; 106 AA.
AC P55248;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Putative protein SH (SH-4).
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NC NCB1_Taxid=10116;
XP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=87149087; PubMed=3547652;
RA Adelstein J.P., Bond C.T., Douglass J., Herbert E.;
RT "Two mammalian genes transcribed from opposite strands of the same
RT DNA locus.";
RT Science 235:1514-1517 (1987).
CC -1- FUNCTION: May be involved with the regulation of GNRH gene
CC expression. It is not known if this protein is transcribed.
CC -1- TISSUE SPECIFICITY: Heart.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
JR EMBL; M15527; AAA42140.1; -.
JR PIR; A29760; A29760.
JR SEQUENCE 106 AA; 11792 MW; P9B8F8DBEABR3E CRC64;

Query Match 29.6%; Score 50; DB 1; Length 106;
Best Local Similarity 47.1%; Pred. No. 7.5;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Y 5 AGAARGASWGMALALL 21
Y 2 AHAVRSKSNWCQTYLL 18

RESULT 10
RT RP1_YEAST STANDARD; PRT; 219 AA.
ID RP1_YEAST STANDARD; PRT; 219 AA.
AC Q05359;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ERP1 protein precursor.
NC ERP1 OR YAR002C-A OR YAR002A.
CC Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
NC NCB1_Taxid=4932;
XP SEQUENCE FROM N.A.
RC STRAIN=8288C / AB972;
RX MEDLINE=95028152; PubMed=7941740;
RA Clark M.W., Keng T., Storms R.K., Zhong W., Fortin N., Zeng B.,
RA Delaney S., Ouellette A.F.F., Barton A.B., Kaback D.B., Bussey H.;
RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
RT the 42 kbp SPO7-CENT-CDC15 region.";
RT Yeast 10:535-541 (1994).
XP CHARACTERIZATION.
XP MEDLINE=99287733; PubMed=10359606;
RA Marzloch M., Henthorn D.C., Herrmann J.M., Wilson R., Thomas D.Y.,
RA Bergeron J.J., Solari R.C., Rowley A.;

```

```

RT "Erp1p and Erp2p, partners for Emp24p and Erv25p in a yeast p24
RT complex.";
RT Mol. Biol. Cell 10:1923-1938 (1999).
CC -1- FUNCTION: Involved in vesicular protein trafficking.
CC -1- SUBUNIT: Associates with EMP24, ERV25 and ERP2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -1- SIMILARITY: Belongs to the EMP24/GP25L family.
CC -1- SIMILARITY: Contains 1 GOLD domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L22015; AAC04958.1; -.
DR GenBank; U00001.2; ERP1.
DR SGP; S0002129; ERP1.
DR GO; GO:0030138; C:COPII-coated vesicle; IDA.
DR GO; GO:0006888; P:ER to Golgi transport; IPI.
DR InterPro; IPR000348; Emp24_SP25L_p24.
DR InterPro; IPR009038; GOLD.
DR Pfam; PF01105; EMP24_GP25L_1.
DR PROSITE; PSS0866; GOLD; 1.
DR Transprot; Protein transport; Transmembrane; Signal;
KM Endoplasmic reticulum.
FT SIGNAL 1 22
FT CHAIN 23 219
FT DOMAIN 23 186
FT TRANSMEM 187 207
FT DOMAIN 208 219
FT DOMAIN 32 131
FT SEQUENCE 219 AA; 24723 MW; 16236C1886FC42E8 CRC64;

Query Match 29.6%; Score 50; DB 1; Length 219;
Best Local Similarity 39.3%; Pred. No. 14;
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 2 RRGAGARGASWGMALALLAVPQM 29
QY 177 RDASEAIVNSRAWMYIVQILVAVTCGM 204

RESULT 11
RT APCC_ECOLI STANDARD; PRT; 514 AA.
ID APCC_ECOLI STANDARD; PRT; 514 AA.
AC P26459;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome bd-II oxidase subunit I (EC 1.10.3.-).
NC APCC OR CYXA OR CBDA OR B0978.
CC Escherichia coli.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
NC NCB1_Taxid=562;
XP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92049231; PubMed=1658595;
RA Dassa J., Feihl H., Marck C., Dion M., Kieffer-Bontemps M.,
RA Boquet P.L.;
RT "A new oxygen-regulated operon in Escherichia coli comprises the
RT genes for a putative third cytochrome oxidase and for pH 2.5 acid
RT phosphatase (appa).";
RT Mol. Gen. Genet. 229:341-352 (1991).
XP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;

```

Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

"The complete genome sequence of *Escherichia coli* K-12.";

Science 277:1453-1474(1997).

[3]

SEQUENCE FROM N.A.

STRAIN=K12;

MEDLINE=97061202; PubMed=8905232;

Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horinouchi T.;

"A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

DNA Res. 3:1137-1155(1996).

[4]

CHARACTERIZATION, AND SEQUENCE OF 496-510.

MEDLINE=96198179; PubMed=8626304;

Sturr M.G., Kulwich T.A., Hicks D.B.;

"Purification of a cytochrome bd terminal oxidase encoded by the *Escherichia coli* app locus from a delta cyo delta cyd strain complemented by genes from *Bacillus firmus* OF4.";

J. Bacteriol. 178:1742-1749(1996).

-1- FUNCTION: CYTOCHROME OXIDASE SUBUNIT.

-1- PATHWAY: Respiratory chain; terminal step.

-1- SUBUNIT: Heterodimer of subunits I and II (probable).

-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Probable).

-1- INDUCTION: IS INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE; ITS SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM AEROBIC TO ANAEROBIC CONDITIONS.

-1- PTM: The N-terminus is blocked.

-1- SIMILARITY: STRONG, TO E.COLI CYD4.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; S63811; AAB20284.1; --

EMBL; AE000019; AAC74063.1; --

EMBL; D90735; BAA35743.1; --

PIR; S17958; S17958.

Ecogene; EGI1380; appC.

InterPro; IPR002585; Bac Ubp Cox.

Pfam; PF01654; Bac Ubp Cox; 1

Oxidoreductase; Electron transport; Transmembrane; Inner membrane; Heme; Complete proteome.

DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).

TRANSSEM 23 42 POTENTIAL.

DOMAIN 43 94 PERIPLASMIC (POTENTIAL).

TRANSSEM 95 114 POTENTIAL.

DOMAIN 115 129 CYTOPLASMIC (POTENTIAL).

TRANSSEM 130 149 POTENTIAL.

DOMAIN 150 187 PERIPLASMIC (POTENTIAL).

TRANSSEM 188 207 POTENTIAL.

DOMAIN 208 219 CYTOPLASMIC (POTENTIAL).

TRANSSEM 220 239 POTENTIAL.

DOMAIN 240 392 PERIPLASMIC (POTENTIAL).

TRANSSEM 393 412 POTENTIAL.

DOMAIN 413 470 CYTOPLASMIC (POTENTIAL).

TRANSSEM 471 490 POTENTIAL.

DOMAIN 491 514 PERIPLASMIC (POTENTIAL).

METAL 186 186 IRON (HEME AXIAL LIGAND).

SEQUENCE 514 AA; 57920 MW; 2D2FBD43429D960D CRC64; (BY SIMILARITY).

Query Match 29.6%; Score 50; DB 1; Length 514;

Best Local Similarity 50.0%; Pred. No. 30;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 16 WALALLMALVAVPGW 29

Db 428 WSLPLFWTAIEAGW 441

-----

RESULT 12

DPM3\_HUMAN STANDARD; PRT; 92 AA.

AC Q9P2X0; O9BXN4; O9BXN5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dolichol-phosphate mannosyltransferase subunit 3 (Dolichol-phosphate mannosyltransferase subunit 3) (Dolichyl-phosphate beta-D-mannosyltransferase subunit 3) (Mannose-P-dolichol synthase subunit 3) (MPP synthase subunit 3) (DPM synthase complex subunit 3) (Prostin 1).

DE DPM3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_Taxid=9606;

[1]

RP SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 1-13, AND CHARACTERIZATION.

RX MEDLINE=20296673; PubMed=10835346;

RA Maeda Y., Tanaka S., Hino J., Kangawa K., Kinoshita T.;

RT "Human dolichol-phosphate-mannose synthase consists of three subunits, DPM1, DPM2 and DPM3.";

RL EMBO J. 19:2475-2482(2000).

[2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RX MEDLINE=21313795; PubMed=11420690;

RA Manos E.J., Kim M.L., Kassir J., Chang P.Y., Wells A., Jones D.A.;

RT "Dolichol-phosphate-mannose-3 (DPM3)/prostin-1 is a novel phospholipase C-gamma regulated gene negatively associated with prostate tumor invasion.";

RL Oncogene 20:2781-2790(2001).

CC -1- FUNCTION: Stabilizer subunit of the dolichol-phosphate-mannose synthase complex.

CC -1- SUBUNIT: Composed of three subunits; DPM1, DPM2 and DPM3.

CC Associated with DPM1 via its C-terminal domain and with DPM2 via its N-terminal portion.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1; Synonyms=Short;

CC IsoId=Q9P2X0-1; Sequence=Displayed;

CC Name=2; Synonyms=Long;

CC IsoId=Q9P2X0-2; Sequence=VSP 001308;

CC -1- SIMILARITY: Belongs to the DPM3 family.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AB028128; BAA96291.1; --

EMBL; AF312922; AAK28487.1; --

EMBL; AF312923; AAK28486.1; --

Genew; HGNC:3007; DPM3.

MTM; 605951; --

DR Transmembrane; Endoplasmic reticulum; Alternative splicing.

KW TRANSSEM 8 28 POTENTIAL.

FT TRANSSEM 37 57 POTENTIAL.

FT VARSPLIC 1 1 M -> MLSVGRLSLVRSFLLRGALLPSLAVTM (in

```

FT FT isoform 2).
FT FT /FTID=VSP_001308.
FT FT V -> L (IN REF. 2).
SQ SEQUENCE 92 AA; 10080 MW; C350B3696842A877 CRC64;

Query Match
Best Local Similarity 45.0%; Score 49.5; DB 1; Length 92;
Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

2y 12 ASWCWALAL---WLAVVG 28
Db 5 AQMLWGLALISGTWALRTG 24

RESULT 13
PEXD CAEEL STANDARD; PRT; 330 AA.
1C Q1951:
2T 15-DEC-1998 (Rel. 37, Created)
3T 28-FEB-2003 (Rel. 41, Last sequence update)
4T 28-FEB-2003 (Rel. 41, Last annotation update)
5E Probable peroxisomal membrane protein PEX13 (Peroxin-13).
6F F32A5.6.
7S Caenorhabditis elegans.
8C Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
9C Rhabditidae; Pelodierinae; Caenorhabditis.
10 NCBI_TaxID=6239;
11 [1]
12 SEQUENCE FROM N.A.
13 STRAIN=Bristol NZ;
14 Pauley A.;
15 Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
16 [2]
17 REVISIONS.
18 Waterston R.;
19 Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
20 -1- FUNCTION: Component of the peroxisomal translocation machinery
21 with PEX14 and PEX17. Functions as a docking factor for the
22 predominantly cytoplasmic PEX1 receptor (PAX10/PEX5) (By
23 similarity).
24 -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (By
25 similarity).
26 -1- SIMILARITY: Contains 1 SH3 domain.
27 -----
28 This SWISS-PROT entry is copyright. It is produced through a collaboration
29 between the Swiss Institute of Bioinformatics and the EMBL outstation -
30 the European Bioinformatics Institute. There are no restrictions on its
31 use by non-profit institutions as long as its content is in no way
32 modified and this statement is not removed. Usage by and for commercial
33 entities requires a license agreement (See http://www.isb-sib.ch/announce/
34 or send an email to license@isb-sib.ch).
35 -----
36 C
37 R EMBL; U00864; AAC4665.2; -.
38 R HSSP; P00519; 1ABU.
39 R WormBep; F32A5.6; CE28290.
40 R InterPro; IPR007223; Peroxin_13_N.
41 R InterPro; IPR001452; SH3.
42 R Pfam; PF04088; Peroxin-13_N; 1.
43 R Pfam; PF00018; SH3; 1.
44 R PRINTS; PR00452; SH3DOMAIN.
45 R ProDom; PD000066; SH3; 1.
46 R SMART; SM00326; SH3; 1.
47 R PROSITE; PS50002; SH3; 1.
48 R Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.
49 T DOMAIN 1 192 LIDENAL (POTENTIAL).
50 T TRANSMEM 193 213 CYTOPLASMIC (POTENTIAL).
51 T DOMAIN 214 330
52 T DOMAIN 235 300 SH3.
53 T SEQUENCE 330 AA; 35765 MW; 7DD647F07A04403 CRC64;

Query Match
Best Local Similarity 44.0%; Score 49.5; DB 1; Length 330;
Matches 11; Conservative 2; Mismatches 11; Indels 1; Gaps 1;

```

```

Qy 6 GARGASWCWALALMLAVVG-W 29
Db 187 GATRPASVNPALFWVAIGPW 211

RESULT 14
INT SYNY3 STANDARD; PRT; 519 AA.
1C P74055;
2T 30-MAY-2000 (Rel. 39, Created)
3T 30-MAY-2000 (Rel. 39, Last sequence update)
4T 28-FEB-2003 (Rel. 41, Last annotation update)
5E Apolipoprotein N-acyltransferase (EC 2.3.1.-) (ALP N-acyltransferase).
6F LNT OR SRO819.
7S Synechocystis sp. (strain PCC 6803).
8C Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
9C NCBI_TaxID=1148;
10 [1]
11 SEQUENCE FROM N.A.
12 MEDLINE=97061201; PubMed=8905231;
13 Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
14 Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
15 Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K., Okumura S.,
16 Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
17 Tabata S.;
18 "Sequence analysis of the genome of the unicellular cyanobacterium
19 Synechocystis sp. strain PCC6803. II. Sequence determination of the
20 entire genome and assignment of potential protein-coding regions.";
21 DNA Res. 3:109-136(1996).
22 -1- FUNCTION: Transfers the fatty acyl group on membrane lipoproteins
23 (By similarity).
24 -1- PATHWAY: Lipoproteins biosynthesis.
25 -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
26 -1- SIMILARITY: Belongs to the apolipoprotein N-acyltransferase
27 family.
28 -----
29 This SWISS-PROT entry is copyright. It is produced through a collaboration
30 between the Swiss Institute of Bioinformatics and the EMBL outstation -
31 the European Bioinformatics Institute. There are no restrictions on its
32 use by non-profit institutions as long as its content is in no way
33 modified and this statement is not removed. Usage by and for commercial
34 entities requires a license agreement (See http://www.isb-sib.ch/announce/
35 or send an email to license@isb-sib.ch).
36 -----
37 C
38 R EMBL; D90911; BAA18131.1; -.
39 R PIR; S75570; S75570.
40 R InterPro; IPR004563; Lnt.
41 R InterPro; IPR003010; Ntlase/CNhydase.
42 R Pfam; PF00795; CN hydrolase; 1.
43 R TIGRfam; TIGR00546; Lnt; 1.
44 DR PROSITE; PS50263; CN HYDROLASE; 1.
45 R Transferase; Acyltransferase; Transmembrane; Complete proteome.
46 FT TRANSMEM 6 26
47 FT TRANSMEM 47 67 POTENTIAL.
48 FT TRANSMEM 83 103 POTENTIAL.
49 FT TRANSMEM 126 146 POTENTIAL.
50 FT TRANSMEM 174 194 POTENTIAL.
51 FT TRANSMEM 206 226 POTENTIAL.
52 FT TRANSMEM 277 297 POTENTIAL.
53 FT TRANSMEM 496 516 POTENTIAL.
54 FT DOMAIN 239 519 CN HYDROLASE.
55 SQ SEQUENCE 519 AA; 58130 MW; A28C75F260E9BBA3 CRC64;

Query Match
Best Local Similarity 38.5%; Score 49.5; DB 1; Length 519;
Matches 10; Conservative 3; Mismatches 2; Indels 11; Gaps 2;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

4 protein - protein search, using sw model

on: June 8, 2004, 13:59:08 ; Search time 5.54167 Seconds

(without alignments)  
1708.068 Million cell updates/sec

File: US-10-010-050a-2\_COPY\_1\_30

Sequence: 1 MRRGAGARGASWCALALMLAVPGWS 30

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_25.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_ivirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	39.1	606	17	Q9HMT4
2	65.5	38.6	205	16	Q8PDE8
3	63.5	37.6	201	16	Q8PDE5
4	62	36.7	588	16	Q82G94
5	59	34.9	198	16	Q8PXY4
6	59	34.9	198	16	Q8PXY4
7	59	34.9	198	16	Q8PXY4
8	59	34.9	198	16	Q8PXY4
9	59	34.9	198	16	Q8PXY4
10	58	34.3	146	11	Q8CEK6
11	57.5	34.0	343	16	Q8G419
12	57	33.7	226	16	Q8DG07
13	57	33.7	366	10	Q84X72
14	56	33.1	214	16	Q84X72
15	56	33.1	485	16	Q88SV3
16	55.5	32.8	371	4	Q96H96

17	55.5	32.8	412	10	Q7XSY0
18	55.5	32.8	695	16	Q8XW44
19	55	32.5	147	10	Q7X917
20	55	32.5	232	16	Q9PEH4
21	55	32.5	232	16	Q87EH4
22	55	32.5	707	10	Q9ZRA6
23	55	32.5	836	16	Q89HK3
24	55	32.5	3019	12	Q92529
25	54.5	32.2	186	10	Q8LNR9
26	54.5	32.2	435	10	Q9XE69
27	54	32.0	54	8	Q9W1A5
28	54	32.0	214	10	Q9XE06
29	54	32.0	225	10	Q84ZK3
30	54	32.0	246	16	Q87AE0
31	54	32.0	666	2	Q8L311
32	53.5	31.7	126	2	Q8WV97
33	53.5	31.7	284	16	Q9PEH9
34	53.5	31.7	284	16	Q87EH8
35	53.5	31.7	749	16	Q82SX8
36	53	31.4	83	9	Q7Y4J5
37	53	31.4	98	17	Q8ZSP9
38	53	31.4	160	16	Q82HM2
39	53	31.4	224	2	Q9EY24
40	53	31.4	428	16	Q8UD14
41	53	31.4	455	16	Q88D16
42	53	31.4	504	16	Q7WVW8
43	52.5	31.1	504	16	Q7WVW4
44	52.5	31.1	220	16	Q8YBF6
45	52.5	31.1	325	16	Q8FWY7

## ALIGNMENTS

RESULT 1  
ID Q9HMT4 PRELIMINARY; PRT; 606 AA.  
AC Q9HMT4;  
DT 01-MAR-2001 (TREMURel. 16, Created)  
DT 01-MAR-2001 (TREMURel. 16, Last sequence update)  
DT 01-JUN-2003 (TREMURel. 24, Last annotation update)  
DE Sulfate transport system permease protein.  
GN CYSY1 OR VNG2396G.  
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahatras G.G., Bergquist B., Pan M., Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroga J., Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A., Leithausen B., Keller K., Cruz R., Danson M.T., Hough D.W., Idenberger T.A., Peck K.F., Pohlschroder M., Spudis C.M., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasasara S.; "Genome sequence of Halobacterium species NRC-1"; Proc. Natl. Acad. Sci. U.S.A. 97:12175-12181(2000).  
CC -!- FUNCTION: PROBABLY PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEMS.  
EMBL: AF005121; AAC20487.1; -.  
PIR: C84390; C84390.  
DR GO: GO:0016021; C: integral to membrane; IEA.  
DR GO: GO:0003700; F: transcription factor activity; IEA.  
DR GO: GO:0005215; F: transporter activity; IEA.  
DR GO: GO:0006355; P: regulation of transcription, DNA-dependent; IEA.  
DR GO: GO:0006810; P: transport; IEA.



```

JR InterPro; IPR000515; BPD_transp.
JR InterPro; IPR000847; HTH_LysR.
JR Pfam; PF00528; BPD_transp; 2.
JR PROSITE; PS00402; BPD_TRANS INN_MEMBER; 1.
JR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
JR Transmembrane; Transport; Complete proteome.
IO SEQUENCE 606 AA; 6306 MW; 256489A20298A28C CRC64;

Query Match      39.1%; Score 66; DB 17; Length 606;
Best Local Similarity 53.8%; Pred. No. 2.1;
Matches 14; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

b      3  RGAGAGRGRASWCALLLMLAVVPG 28
      |||:|||||:|:|:|:|:|
508  RALGASGRRAIMVDELPLVWGVVAG 533

ESTUT 2
C QBPDP8      PRELIMINARY;      PRT;      205 AA.
T 01-OCT-2002 (TREMBlrel. 22, Created)
T 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
T 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
E Hypothetical protein XCC0380.
N XCC0380.
C Xanthomonas campestris (pv. campestris).
C Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
C Xanthomonadaceae; Xanthomonas.
N NCBI_TaxID=340;
[1]
P SEQUENCE FROM N.A.
C STRAIN=ATCC 33913 / NCPPB 528;
X MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarote G., Camavan F., Cardozo J., Chamberg F., Clapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitejima J.P.,
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
N Nature 417:459-463(2002).
L Nature 417:459-463(2002).
3 EMBL; AE012134; AAM39699.1; -.
3 InterPro; IPR003848; DUF218.
3 Pfam; PR02698; DUF218; 1.
4 Hypothetical protein; Complete proteome.
2 SEQUENCE 205 AA; 23048 MW; CCC1D7C8BAF9A482 CRC64;

Query Match      38.8%; Score 65.5; DB 16; Length 205;
Best Local Similarity 42.4%; Pred. No. 0.92;
Matches 14; Conservative 3; Mismatches 11; Indels 5; Gaps 2;

/      2  RRGAAGRGRASWCMA-----LALLMLAVVPGM 29
      |||:|||||:|:|:|:|:|
3  RRVSRPRGRLGIMGWGRMLCMALIMLVAGM 35

ESTUT 3
C QBPDP8      PRELIMINARY;      PRT;      201 AA.
T 01-OCT-2002 (TREMBlrel. 22, Created)
T 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
T 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

```

```

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein XAC0380.
GN XAC0380.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
[1]
P SEQUENCE FROM N.A.
C STRAIN=306 / ATCC 13902 / XV 101;
X MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarote G., Camavan F., Cardozo J., Chamberg F., Clapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitejima J.P.,
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
N Nature 417:459-463(2002).
L Nature 417:459-463(2002).
3 EMBL; AE011664; AAM35272.1; -.
3 InterPro; IPR003848; DUF218.
3 Pfam; PR02698; DUF218; 1.
4 Hypothetical protein; Complete proteome.
2 SEQUENCE 201 AA; 22724 MW; 62D3709290F3D458 CRC64;

Query Match      37.6%; Score 63.5; DB 16; Length 201;
Best Local Similarity 44.4%; Pred. No. 1.6;
Matches 12; Conservative 2; Mismatches 8; Indels 5; Gaps 1;

CY      8  AGRASWCMA-----LALLMLAVVPGM 29
      |||:|||||:|:|:|:|:|
DB      5  SRGLGIMGWGRMLCMALIMLVAVAGM 31

RESULT 4
AC 082694      PRELIMINARY;      PRT;      588 AA.
ID 082694;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV4004.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
[1]
P SEQUENCE FROM N.A.
C STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
X MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
P SEQUENCE FROM N.A.
C STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
X MEDLINE=22608306; PubMed=12692562;
Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

```

A Sakaki Y., Hattori M., Omura S.;  
T "Complete genome sequence and comparative analysis of the industrial  
L microorganism *Streptomyces avermitilis*."; Nat. Biotechnol. 21:526-531(2003).  
R EMBL; AP005037; BAC71716.1; -  
W Hypothetical protein; Complete proteome.  
M SEQUENCE 588 AA; 63428 MW; D8DC5F3BE49A321 CRC64;

Query Match 36.7%; Score 62; DB 16; Length 588;  
Best Local Similarity 28.6%; Pred. No. 6.7;  
Matches 14; Conservative 5; Mismatches 8; Indels 22; Gaps 1;

3 GRASMCWALALMLAVVPGW 29  
167 RGAGSRTLVATKLTLLIGVGRYGDKSMWAGLNLAPARV 215

RESULT 5  
PRELIMINARY; PRT; 198 AA.

08P1Y4; 01-OCT-2002 (TREMBlrel. 22, Created)  
01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
TonB-like protein.  
XAC1651.  
Xanthomonas axonopodis (pv. citri).  
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
Xanthomonadaceae; Xanthomonas.  
NCBI\_TaxID=92829;

SEQUENCE FROM N.A.

STRAIN=306 / ATCC 13902 / XV 101;  
MEDLINE=22022145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
Camarote G., Cannavan F., Cardozo J., Chamberg F., Clapina L.P.,  
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
Seubal J.C., Kitajima J.P.;  
"Comparison of the genomes of two Xanthomonas pathogens with differing  
host specificities."; Nature 417:459-463(2002).  
EMBL; AE011797; AAM36519.1; -  
InterPro; IPR006260; TonB\_C.  
TIGRFAMs; TIGR01352; TonB\_Cterm; 1.  
Complete proteome.  
SEQUENCE 198 AA; 21748 MW; FB9DF173EDC52961 CRC64;

Query Match 34.9%; Score 59; DB 16; Length 198;  
Best Local Similarity 43.5%; Pred. No. 6.1;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

7 AARGASMCWALALMLAVVPGW 29  
36 ASDGASPMWATVAVLVVPSW 58

PRELIMINARY; PRT; 198 AA.

08P1Y4; 01-OCT-2002 (TREMBlrel. 22, Created)  
01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE TonB-like protein.  
GN XCC1592.

OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
NCBI\_TaxID=340;

SEQUENCE FROM N.A.  
STRAIN=ATCC 33913 / NCPPB 528;  
MEDLINE=22022145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
Camarote G., Cannavan F., Cardozo J., Chamberg F., Clapina L.P.,  
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
Seubal J.C., Kitajima J.P.;  
"Comparison of the genomes of two Xanthomonas pathogens with differing  
host specificities."; Nature 417:459-463(2002).  
EMBL; AE012260; AAM40887.1; -  
InterPro; IPR006260; TonB\_C.  
TIGRFAMs; TIGR01352; TonB\_Cterm; 1.  
Complete proteome.  
SEQUENCE 198 AA; 21744 MW; 78A3626C6142C60C CRC64;

Query Match 34.9%; Score 59; DB 16; Length 198;  
Best Local Similarity 50.0%; Pred. No. 6.1;  
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

10 GRASMCWALALMLAVVPGW 29  
41 GASPMWATVAVLVVPSW 60

PRELIMINARY; PRT; 304 AA.

01-OCT-2003 (TREMBlrel. 25, Created)  
01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
Putative membrane protein.  
BN4263.  
Bordetella bronchiseptica (Alcaligenes bronchiseptica).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
NCBI\_TaxID=518;

SEQUENCE FROM N.A.  
STRAIN=RB50 / ATCC BAA-588;  
MEDLINE=22827954; PubMed=12910271;  
Parkhill J., Sebatian M., Preston A., Murphy L.D., Thomson N.,  
Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
Achtmann M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jorgels K.,  
Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,  
Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
Sharp S., Simmonds M., Skelton J., Squares R., Squares K., Stevens K.,  
Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;  
"Comparative analysis of the genome sequences of Bordetella pertussis,  
Bordetella parapertussis and Bordetella bronchiseptica.";

```

L1  Nat. Genet. 35:32-40 (2003).
L2  EMBL; BX640450; CAE34627.1; -.
L3  Complete proteome.
L4  SEQUENCE 304 AA; 31640 MW; 7AF46DA2FB95EFC0 CRC64;
L5  Query Match 34.9%; Score 59; DB 16; Length 304;
L6  Best Local Similarity 39.5%; Pred. No. 9;
L7  Matches 15; Conservative 2; Mismatches 9; Indels 12; Gaps 2;

L8  4 GAGAGRGRASW---CWAL-----ALLMLAVVPGW 29
L9  182 GARASRTLGGRITCWALVSVPALPRAVLMLADAAW 219

L10  RESULT 8
L11  17M451
L12  D 07M451 PRELIMINARY; PRT; 304 AA.
L13  C 07M451;
L14  T 01-OCT-2003 (TREMBlrel. 25, Created)
L15  T 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
L16  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L17  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L18  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L19  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L20  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L21  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L22  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L23  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L24  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L25  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L26  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L27  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L28  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L29  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L30  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L31  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L32  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L33  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L34  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L35  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L36  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L37  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L38  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L39  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L40  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L41  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L42  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L43  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L44  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L45  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L46  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L47  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L48  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L49  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L50  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L51  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L52  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L53  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L54  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L55  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L56  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L57  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L58  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L59  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L60  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L61  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L62  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L63  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L64  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L65  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L66  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L67  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L68  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L69  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L70  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L71  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L72  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L73  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L74  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L75  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L76  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L77  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L78  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L79  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L80  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L81  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L82  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L83  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L84  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L85  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L86  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L87  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L88  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L89  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L90  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L91  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L92  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L93  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L94  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L95  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L96  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L97  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L98  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L99  T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
L100 T 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

```

```

RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481411;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
DR EMBL; AF603644; CAC49196.1; -.
DR PIR; D95941; D95941.
DR GO; GO:0046821; Cytochrome c heme B5.
DR GO; GO:0005489; Electron transport activity; IEA.
DR GO; GO:0006118; Electron transport; IEA.
DR InterPro; IPR003345; Cytochrome b5.
DR PROSITE; PS00190; CYTOCHROME_C_1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 412 AA; 46126 MW; 8088B8E0C6F6838 CRC64;

QY 7 AARGRASCWA-----LALLMLAVVP 27
DB 276 ANKGRPTWTLATVLFIAIMWSTVP 302

RESULT 10
Q8CEK6
ID Q8CEK6 PRELIMINARY; PRT; 146 AA.
AC Q8CEK6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN 9330132005RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dienecephalon;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK020369; BAC25627.1; -.
DR MGD; MGI:1924802; 9330132005RIK.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 146 AA; 16423 MW; AF4B10EB630CB6D5 CRC64;

QY 2 RRGAGRGRASW---CWALALIM 22
DB 44 RGPAGRRSQSSCWCLGLW 64

RESULT 11
Q8G419
ID Q8G419 PRELIMINARY; PRT; 343 AA.
AC Q8G419;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN BL1576.

```



```

2N CLS OR LP 3273.
2S Lactobacillus plantarum.
2C Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
2X Lactobacillus.
2Y NCBI_TaxID=1590;
2Z [1]
2P SEQUENCE FROM N.A.
2C STRAIN=NCIMB 8826 / WCFS1;
2X MEDLINE=22480296; PubMed=12566566;
2A Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
2A Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
2A Eilers M.W.E.J., Silekema W., Klein Lankhorst R.M., Bron P.A.,
2A Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
2A De Vos W.M., Siezen R.J.;
2T "Complete genome sequence of Lactobacillus plantarum WCFS1.";
2L Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
2R EMBL; AL935261; CAD65406.1; -.
2R GO; GO:0016740; F:transferase activity; IEA.
2R GO; GO:0008152; P:metabolism; IEA.
2R InterPro: IPR001736; PLD.
2R Pfam; PF00614; PLDc; 2.
2R PROSITE; PS50035; PLD; 2.
2M Transferase; Complete proteome.
3Q SEQUENCE 485 AA; 55778 MW; C725C28841DF8D31 CRC64;

```

Query Match 33.1%; Score 56; DB 16; Length 485;  
 Best Local Similarity 46.7%; Pred. No. 33;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

2Y 12 ASWCNALALMLAV 26
2b 36 ATWAMLVITWLPPI 50

```

Search completed: June 8, 2004, 14:07:06  
 Job time : 5.54167 secs



reagents to detect zsig46 or cells expressing it, e.g. for assessing thyroid function to produce anti-idiotypic antibodies, for affinity purification of zsig46, to screen expression libraries, to neutralise zsig46 activity, and to deliver toxins, radioisotopes etc. for therapeutic or diagnostic purposes. Agonists of the product can be used to promote growth, differentiation and proliferation of specific cell types, e.g. for treating (extra)thyroid diseases or as additive to cell cultures

Sequence 346 AA:

Query Match 100.0%; Score 1758; DB 2; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-169;  
 Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCPTGSPPIVMEGDDIEVFRLOAPVW 60  
 b RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCPTGSPPIVMEGDDIEVFRLOAPVW 90  
 y 61 EFXYGDLIGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCPTFPHLRPMDAPFMCNOG 120  
 b 91 EFXYGDLIGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCPTFPHLRPMDAPFMCNOG 150  
 y 121 AACFEESIDVHWKENTLVQVATISGMNOMAKVYKQDNETGIYETWNVKASPEKGA 180  
 b 151 AACFEESIDVHWKENTLVQVATISGMNOMAKVYKQDNETGIYETWNVKASPEKGA 210  
 y 181 ETWFDSDYSCSFVLRTEFNKLAEGAFENKNIETNYTRIFLYSGEPTLYGNETSVEGPTGNK 240  
 b 211 ETWFDSDYSCSFVLRTEFNKLAEGAFENKNIETNYTRIFLYSGEPTLYGNETSVEGPTGNK 270  
 y 241 TLGLAIKRFYYPFKPHLPTEKFLSLLOIFDAVIVHKQFLFYNEFWELPMKPPFIKIT 300  
 b 271 TLGLAIKRFYYPFKPHLPTEKFLSLLOIFDAVIVHKQFLFYNEFWELPMKPPFIKIT 330  
 y 301 YEEIPLPIRNKTLISGL 316  
 b 331 YEEIPLPIRNKTLISGL 346

RESULT 2  
 AA41323  
 D AA41323 standard; protein; 347 AA.  
 C AA41323;  
 X  
 T 02-DEC-1999 (first entry)  
 E Human secreted protein encoded by gene 16 clone HMZAD77.  
 K Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 M diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 M developmental abnormality; foetal deficiency; blood; allergy; renal;  
 M immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 M inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 M cognitive disorder; schizophrenia; prostate; obesity; osteoclast; chymus;  
 M osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 M endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 K Homo sapiens.  
 K MO947540-A1.  
 X  
 Y 23-SEP-1999.  
 C  
 F 18-MAR-1999; 99WO-US005804.  
 X  
 Z 19-MAR-1998; 98US-0078563P.  
 Z 19-MAR-1998; 98US-0078566P.  
 Z 19-MAR-1998; 98US-0078573P.  
 Z 19-MAR-1998; 98US-0078574P.  
 Z 19-MAR-1998; 98US-0078576P.  
 Z 19-MAR-1998; 98US-0078577P.

PR 19-MAR-1998; 98US-0078578P.  
 PR 19-MAR-1998; 98US-0078579P.  
 PR 19-MAR-1998; 98US-0078581P.  
 PR 01-APR-1998; 98US-0080312P.  
 PR 01-APR-1998; 98US-0080313P.  
 PR 01-APR-1998; 98US-0080314P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;  
 PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
 PI Shi Y, Moore PA;  
 DR WPI: 1999-562050/47.  
 DR N-PSDB; AA224826.  
 PT New isolated human genes, useful for diagnosis and treatment of e.g.  
 PT cancers, neurological disorders, immune diseases, inflammation or blood  
 disorders.  
 PS Claim 11; Page 367-368; 484pp; English.  
 XX  
 CC This sequence represents a secreted human protein encoded by the gene  
 CC clone detailed in the descriptor line. The gene can be used to generate  
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
 CC portion (e.g. AA224802) for increasing the stability of the fused protein  
 CC as compared to the human protein only. The invention relates to 95 novel  
 CC genes and their fragments (nucleic acid sequences: AA224811-224907; amino  
 CC acid sequences AA41308-Y41404) which are useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
 CC pathological conditions can be diagnosed by determining the amount of the  
 CC new polypeptides in a sample or by determining the presence of mutations  
 CC in the new polynucleotides. Specific uses are described for each of the  
 CC 95 polynucleotides, based on which tissues they are most highly expressed  
 CC in (see AA224811 for described uses)  
 XX  
 SQ Sequence 347 AA;

Query Match 99.4%; Score 1747; DB 2; Length 347;  
 Best Local Similarity 99.4%; Pred. No. 4.7e-168;  
 Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCPTGSPPIVMEGDDIEVFRLOAPVW 60  
 Db 31 RVSGIPSRHRMPVYKRFDPKPPYCOAKYTPCPTGSPPIVMEGDDIEVFRLOAPVW 90  
 Qy 61 EFXYGDLIGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCPTFPHLRPMDAPFMCNOG 120  
 Db 91 EFXYGDLIGHLKIMHDAIGFRSTLTGKNYTWEMWELFOLGNCPTFPHLRPMDAPFMCNOG 150  
 Qy 121 AACFEESIDVHWKENTLVQVATISGMNOMAKVYKQDNETGIYETWNVKASPEKGA 180  
 Db 151 AACFEESIDVHWKENTLVQVATISGMNOMAKVYKQDNETGIYETWNVKASPEKGA 210  
 Qy 181 ETWFDSDYSCSFVLRTEFNKLAEGAFENKNIETNYTRIFLYSGEPTLYGNETSVEGPTGNK 240  
 Db 211 ETWFDSDYSCSFVLRTEFNKLAEGAFENKNIETNYTRIFLYSGEPTLYGNETSVEGPTGNK 270  
 Qy 241 TLGLAIKRFYYPFKPHLPTEKFLSLLOIFDAVIVHKQFLFYNEFWELPMKPPFIKIT 300  
 Db 271 TLGLAIKRFYYPFKPHLPTEKFLSLLOIFDAVIVHKQFLFYNEFWELPMKPPFIKIT 330  
 Qy 301 YEEIPLPIRNKTLISGL 316  
 Db 331 YEEIPLPIRNKTLISGL 346

RESULT 3  
 AAM93870  
 ID AAM93870 standard; protein; 358 AA.  
 AC AAM93870;  
 XX

06-NOV-2001 (first entry)  
 Human polypeptide, SEQ ID NO: 3978.  
 Human; full length cDNA; cDNA synthesis; oligo-capping.  
 Homo sapiens.  
 EP130094-A2.  
 05-SEP-2001.  
 07-JUL-2000; 2000BP-00114089.  
 08-JUL-1999; 99JP-00194486.  
 11-JAN-2000; 2000JP-00118774.  
 02-MAY-2000; 2000JP-00183765.  
 (HELI-) HELIX RES INST.  
 Ota T, Nishikawa T, Isogai T, Hayashi K, Iehi S, Kawai Y;  
 Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 WPI; 2001-524255/58.  
 N-PSDB; AAK94829.  
 830 Primers useful for synthesizing full length cDNA clones and their use  
 in genetic manipulation.  
 Claim 8; SEQ ID NO 3978; 1380bp + Sequence Listing; English.  
 The invention relates to primers for synthesizing full length cDNA  
 clones. 830 cDNA molecules encoding a human protein have been isolated  
 and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
 been determined. Primers for synthesizing the full length cDNA are useful  
 for clarifying the function of the protein encoded by the cDNA. The full  
 length clones were obtained by construction of full length enriched cDNA  
 libraries that were synthesised by the oligo-capping method. The primers  
 enable the production of the full length cDNA easily without any special  
 methods. The present sequence is a polypeptide encoded by a full length  
 human cDNA of the invention. Note: The sequence data for this patent did  
 not form part of the printed specification, but was obtained in CD-ROM  
 format directly from EPO

Sequence 358 AA;  
 Query Match 99.2%; Score 1744; DB 4; Length 358;  
 Best local Similarity 99.4%; Pred. No. 9.9e-168;  
 Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 1 RVSGIPSRHNPVYKRRFDPKPPYCOAKYTCPTGSPPIPVWEGDDIEVFLQAPVW 60  
 43 RVSGIPSRHNPVYKRRFDPKPPYCOAKYTCPTGSPPIPVWEGDDIEVFLQAPVW 102  
 61 EFKYGDLLGHLKIHDAIGRRSTLTGKNYTWEMVFLQNGCTFPHLRPEMDAPFWCNOG 120  
 103 EFKYGDLLGHLKIHDAIGRRSTLTGKNYTWEMVFLQNGCTFPHLRPEMDAPFWCNOG 162  
 121 AACFBSGIDVHMKENGTLVQVATISGMENQAKWKQKONETGIYYETWNVKASPEKGA 180  
 163 AACFBSGIDVHMKENGTLVQVATISGMENQAKWKQKONETGIYYETWNVKASPEKGA 222  
 181 EFWDSYDCKSFVLRNTKLAEGFAFKNIETNTRIIFVSGEPTYVGNESVAGPTGNK 240  
 223 EFWDSYDCKSFVLRNTKLAEGFAFKNIETNTRIIFVSGEPTYVGNESVAGPTGNK 282  
 241 TLGLAIRFYPPKPHLPYKEFLISLLQIFDAVIVHKQFYLFTNFEYFLPMKFPPIKIT 300  
 283 TLGLAIRFYPPKPHLPYKEFLISLLQIFDAVIVHKQFYLFTNFEYFLPMKFPPIKIT 342  
 301 YEELPIPIRKKTLSGL 316  
 343 YEELPIPIRKKTLSGL 358

RESULT 4  
 AAM19604  
 ID AAM19604 standard; protein; 163 AA.  
 XX  
 AC AAM19604;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Peptide #6038 encoded by probe for measuring cervical gene expression.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX  
 KM cervical cancer.  
 OS  
 XX Homo sapiens.  
 EN  
 XX MO200157278-A2.  
 PD  
 XX 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US000670.  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human cervical epithelial cells.  
 PS  
 XX Claim 27; SEQ ID NO 24430; 487bp; English.  
 CC  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENPs; see A110068-A1129459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 163 AA;  
 Query Match 50.4%; Score 886; DB 4; Length 163;  
 Best local Similarity 100.0%; Pred. No. 2.8e-81;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 154 AKWKQDNETGIYYETWNVKASPEKGAETWFDSDYDCKFVLRNTKLAEGFAFKNIETN 213  
 DB 1 AKWKQDNETGIYYETWNVKASPEKGAETWFDSDYDCKFVLRNTKLAEGFAFKNIETN 60  
 QY 214 YTRIFLYSGEPTYVGNESVAGPTGNKTLGLAIRFYPPKPHLPYKEFLISLLQIFDAV 273  
 DB 61 YTRIFLYSGEPTYVGNESVAGPTGNKTLGLAIRFYPPKPHLPYKEFLISLLQIFDAV 120  
 QY 274 IYHKQFYLFTNFEYFLPMKFPPIKITYERIPPIRKKTLSGL 316  
 DB 121 IYHKQFYLFTNFEYFLPMKFPPIKITYERIPPIRKKTLSGL 163

RESULT 5



BB39239  
D ABB39239 standard; peptide; 163 AA.  
X  
X ABB39239;  
X  
X 04-FEB-2002 (first entry)  
T  
X Peptide #6745 encoded by human foetal liver single exon probe.  
X  
X Human; foetal liver; gene expression; single exon nucleic acid probe.  
S Homo sapiens.  
N WO200157272-A2.  
X  
X 09-AUG-2001.  
D  
X  
X 30-JAN-2001; 2001WO-US000669.  
F  
X  
X 04-FEB-2000; 2000US-0180312P.  
R 26-MAY-2000; 2000US-0207456P.  
R 30-JUN-2000; 2000US-00608408.  
R 03-AUG-2000; 2000US-00632366.  
R 21-SEP-2000; 2000US-0234687P.  
R 27-SEP-2000; 2000US-0236359P.  
R 04-OCT-2000; 2000GB-00024263.  
X  
X (MOLE-) MOLECULAR DYNAMICS INC.  
X  
X Penn SG, Hanzel DK, Chen W, Rank DR;  
X WPI; 2001-483447/52.  
X  
X Human genome-derived single exon nucleic acid probes useful for analyzing  
T gene expression in human fetal liver.  
X  
X Claim 27; SEQ ID NO 31874; 639pp + Sequence listing; English.  
S  
X The invention relates to a single exon nucleic acid probe for measuring  
C human gene expression in a sample derived from human foetal liver. The  
C single exon nucleic acid probes may be used for predicting, measuring and  
C displaying gene expression in samples derived from human fetal liver. The  
C present sequence is a peptide encoded by a single exon nucleic acid probe  
C of the invention. Note: The sequence data for this patent did not form  
C part of the printed specification, but was obtained in electronic format  
C directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
X  
X Sequence 163 AA;  
Q  
Query Match 50.4%; Score 886; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.8e-81;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Y 154 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEGAFKNIETN 213  
b 1 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEGAFKNIETN 60  
Y 214 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKPEFLSLLOTFDAY 273  
b 61 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKPEFLSLLOTFDAY 123  
Y 274 IVHKQFYLFFNFYFWFLPMKFPFIKITVEEIPLPINRKTLSGL 316  
b 121 IVHKQFYLFFNFYFWFLPMKFPFIKITVEEIPLPINRKTLSGL 163  
RESULT 6  
D ABB32740 standard; protein; 163 AA.  
X  
X AAB32740;  
X  
X 17-OCT-2001 (first entry)  
T

XX  
DE Peptide #6777 encoded by probe for measuring placental gene expression.  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
XW genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200157272-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US000663.  
PF  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-48897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
PT  
X  
X Claim 27; SEQ ID NO 33009; 654pp; English.  
S  
X The present invention relates to single exon nucleic acid probes (SENP:  
CC see A013135-A0157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
X Sequence 163 AA;  
SQ  
Query Match 50.4%; Score 886; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.8e-81;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 154 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEGAFKNIETN 213  
Db 1 AKWVKODNETGIYETWNVKASPEKGAETWFDSDSKFVLRFTFNKLAEGAFKNIETN 60  
QY 214 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKPEFLSLLOTFDAY 273  
Db 61 YTRIFLYSGEPYTLGNETSVPPTGNKKTGLAIKRFYFPKPHLPKPEFLSLLOTFDAY 120  
QY 274 IVHKQFYLFFNFYFWFLPMKFPFIKITVEEIPLPINRKTLSGL 316  
Db 121 IVHKQFYLFFNFYFWFLPMKFPFIKITVEEIPLPINRKTLSGL 163  
RESULT 7  
ID ABB24080 standard; protein; 163 AA.  
XX  
XX ABB24080;  
XX  
XX 23-JAN-2002 (first entry)  
DT  
XX  
XX Protein #6079 encoded by probe for measuring heart cell gene expression.  
DE  
XX Human; gene expression; heart; microarray; vascular system;  
XW cardiovascular disease; hypertension; cardiac arrhythmia;  
XW congenital heart disease.  
XX

Homo sapiens.

W0200157274-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000666.

04-FEB-2000; 2000US-0180312P.  
 26-MAY-2000; 2000US-0207456P.  
 30-JUN-2000; 2000US-00608408.  
 03-AUG-2000; 2000US-00632366.  
 21-SEP-2000; 2000US-0234687P.  
 27-SEP-2000; 2000US-0236359P.  
 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488899/53.

Single exon nucleic acid probes for analyzing gene expression in human hearts.

Claim 15; SEQ ID NO 25850; 530bp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 163 AA:

Query Match 50.4%; Score 886; DB 4; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-81;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

154 AKWVKDNEGTGIYETWNVKASPEKGAETWDSYDCKFVLRFTFNKLAERGAERKNIETN 213  
 1 AKWVKDNEGTGIYETWNVKASPEKGAETWDSYDCKFVLRFTFNKLAERGAERKNIETN 60  
 214 YTRIFLYSGEPTVLGNSTVFGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAV 273  
 61 YTRIFLYSGEPTVLGNSTVFGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAV 120  
 274 IVHKQFYLFYNFPEYWFLEPMKFPPIKITYEELPIPIRNKTLISGL 316  
 121 IVHKQFYLFYNFPEYWFLEPMKFPPIKITYEELPIPIRNKTLISGL 163

SHUT 8  
 WT2492

AAM72492 standard; protein; 163 AA.

AAM72492;

06-NOV-2001 (first entry)

Human bone marrow expressed probe encoded protein SEQ ID NO: 32738.

Human; bone marrow expressed exon; gene expression analysis; probe; microarray; cancer; leukaemia; lymphoma; myeloma.

Homo sapiens.

XX  
 PN W0200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX

30-JAN-2001; 2001WO-US000666.

XX  
 PF 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.

Example 4; SEQ ID NO 32738; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention

Sequence 163 AA:

Query Match 50.4%; Score 886; DB 4; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-81;  
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 AKWVKDNEGTGIYETWNVKASPEKGAETWDSYDCKFVLRFTFNKLAERGAERKNIETN 213  
 DB 1 AKWVKDNEGTGIYETWNVKASPEKGAETWDSYDCKFVLRFTFNKLAERGAERKNIETN 60  
 QY 214 YTRIFLYSGEPTVLGNSTVFGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAV 273  
 DB 61 YTRIFLYSGEPTVLGNSTVFGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAV 120  
 QY 274 IVHKQFYLFYNFPEYWFLEPMKFPPIKITYEELPIPIRNKTLISGL 316  
 DB 121 IVHKQFYLFYNFPEYWFLEPMKFPPIKITYEELPIPIRNKTLISGL 163

RESULT 9  
 AAM59901

ID AAM59901 standard; protein; 163 AA.

AAM59901;

05-NOV-2001 (first entry)

Human brain expressed single exon probe encoded protein SEQ ID NO: 32006.

Human; brain expressed exon; gene expression analysis; probe; microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

Homo sapiens.

W0200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000667.

Query Match	Best Local Similarity	Score	DB #	Length	Matches	Conservative	Mismatches	Indels	Gaps
154 AKWVQDNDETGIYETWVKASPEKGAETWEDSDSKFYLRTFNMLAEFGAEFNIETN	100.0%	86	4	163	163	0	0	0	0
1 AKWQKQDNDETGIYETWVKASPEKGAETWEDSDSKFYLRTFNMLAEFGAEFNIETN	100.0%	86	4	163	163	0	0	0	0
214 YTRIFLYSGEPTVYGNETSVAFGPTGNKTGLAIKRFYYPKPHLPKKEFLLSLQIFDAV	100.0%	86	4	163	163	0	0	0	0
61 YTRIFLYSGEPTVYGNETSVAFGPTGNKTGLAIKRFYYPKPHLPKKEFLLSLQIFDAV	100.0%	86	4	163	163	0	0	0	0
274 IVHKQFYLFEVFEWYFPLPMKPEPKLTYEELPIPKNTLSGL	100.0%	86	4	163	163	0	0	0	0
121 IVHKQFYLFEVFEWYFPLPMKPEPKLTYEELPIPKNTLSGL	100.0%	86	4	163	163	0	0	0	0

PR	04-OCT-2000; 2000GB-00024263.
XX	(MOE-) MOLECULAR DYNAMICS INC.
PA	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-48898/53.
PI	
XX	
DR	Human genome-derived single exon nucleic acid probes useful for analyzing
XX	gene expression in human adult liver.
PT	
PS	Claim 27; SEQ ID NO 32831; 658bp; English.
XX	
CC	The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC	measuring human gene expression in a sample derived from human adult
CC	liver, comprising one of 13109 defined nucleotide sequences given in the
CC	specification (or complements/ fragments). The probe hybridises at high
CC	stringency to a nucleic acid molecule expressed in the human adult liver.
CC	(I) may be used for predicting, measuring and displaying gene expression
CC	in samples derived from human adult liver. The genes identified may be
CC	involved in genetic liver diseases such as cirrhosis,
CC	hyperlipoproteinemia, hyperlipidemia and hypercholesterolaemia which is
CC	associated with coronary heart disease. ABG47348-ABG59930 represent human
CC	liver single exon encoded peptides of the invention. Note: The sequence
CC	information for this patent does not appear in the printed specification
CC	but was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
CC	
SQ	Sequence 163 AA;
	Query Match                      50.4%; Score 886; DB 4; Length 163;
	Best Local Similarity    100.0%; Pred. No. 2.8e-81;
	Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	
DB	154 AKWKQNDNETGIYYETNMVKSPEKGATWPDSDSCSFVLTEFKALFEFGAFPKNIEN 213   1 AKWKQNDNETGIYYETNMVKSPEKGATWPDSDSCSFVLTEFKALFEFGAFPKNIEN 60
QY	214 YTRIFLVSGETTYIGNETSVPFGTGNKTLGLAIRKYFPFPHLPTEKEFLSLLOIPDAV 273   61 YTRIFLVSGETTYIGNETSVPFGTGNKTLGLAIRKYFPFPHLPTEKEFLSLLOIPDAV 120
DB	274 IVHQPFILFNFFEFWFLPMKPFFIKITYEEIPLPIRNKTLSGL 316   121 IVHQPFILFNFFEFWFLPMKPFFIKITYEEIPLPIRNKTLSGL 163
RESULT 11	
ABG42312	
XX	ABG42312 standard; peptide: 163 AA.
XX	
XX	ABG42312;
DT	19-AUG-2002 (first entry)
XX	
DE	Human peptide encoded by genome-derived single exon probe SEQ ID 31977.
XX	
KM	Human; single exon probe; asthma; lung cancer; COPD; ILD;
KM	chronic obstructive pulmonary disease; interstitial lung disease;
KM	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KM	tubercous sclerosis; Gaucher's disease; Niemann-Pick disease;
KM	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KM	pulmonary histiocytosis; lymphangioleiomyomatosis; Kartagener syndrome;
KM	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KM	primary ciliary dyskinesia; pulmonary hypertension;
XX	hyaline membrane disease.
OS	Homo sapiens.
XX	
FN	WO200186003--A2.
XX	
ID	15-NOV-2001.

PR	04-OCT-2000; 2000GB-00024263.
XX	(MOE-) MOLECULAR DYNAMICS INC.
PA	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-48898/53.
PI	
XX	
DR	Human genome-derived single exon nucleic acid probes useful for analyzing
XX	gene expression in human adult liver.
PT	
PS	Claim 27; SEQ ID NO 32831; 658bp; English.
XX	
CC	The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC	measuring human gene expression in a sample derived from human adult
CC	liver, comprising one of 13109 defined nucleotide sequences given in the
CC	specification (or complements/ fragments). The probe hybridises at high
CC	stringency to a nucleic acid molecule expressed in the human adult liver.
CC	(I) may be used for predicting, measuring and displaying gene expression
CC	in samples derived from human adult liver. The genes identified may be
CC	involved in genetic liver diseases such as cirrhosis,
CC	hyperlipoproteinemia, hyperlipidemia and hypercholesterolaemia which is
CC	associated with coronary heart disease. ABG47348-ABG59930 represent human
CC	liver single exon encoded peptides of the invention. Note: The sequence
CC	information for this patent does not appear in the printed specification
CC	but was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
CC	
SQ	Sequence 163 AA;
	Query Match                      50.4%; Score 886; DB 4; Length 163;
	Best Local Similarity        100.0%; Pred. No. 2.8e-81;
	Matches 163; Conservative    0; Mismatches 0; Indels    0; Gaps     0
QY	
DB	154 AKWKQNDNETGIYYETNMVKSPEKGATWDPSYDCSKFVLTEYNKAEFGAFNRIEN 213   1 AKWKQNDNETGIYYETNMVKSPEKGATWDPSYDCSKFVLTEYNKAEFGAFNRIEN 60
QY	214 YTRIFLVSGETTYIGNETSVPFGTGNKTGLAIRKYFPFKPHLPTEKEFLSLIQIPDAV 273   61 YTRIFLVSGETTYIGNETSVPFGTGNKTGLAIRKYFPFKPHLPTEKEFLSLIQIPDAV 120
DB	274 IVHQFYLFNFYEYFWFLPMKPPFIKITIYEELPLPINKTLISGL 316   121 IVHQFYLFNFYEYFWFLPMKPPFIKITIYEELPLPINKTLISGL 163
RESULT 11	
ABG42312	
XX	ABG42312 standard; peptide: 163 AA.
XX	
XX	ABG42312;
DT	19-AUG-2002 (first entry)
XX	
DE	Human peptide encoded by genome-derived single exon probe SEQ ID 31977.
XX	
KM	Human; single exon probe; asthma; lung cancer; COPD; ILD;
KM	chronic obstructive pulmonary disease; interstitial lung disease;
KM	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KM	tubercous sclerosis; Gaucher's disease; Niemann-Pick disease;
KM	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KM	pulmonary histiocytosis; lymphangioleiomyomatosis; Kartagener syndrome;
KM	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KM	primary ciliary dyskinesia; pulmonary hypertension; hyaline membrane disease.
OS	Homo sapiens.
FN	WO200186003--A2.
XX	
XX	15-NOV-2001.



CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence data represents a human ovarian antigen of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 89 AA;

Query Match 26.8%; Score 472; DB 5; Length 89;  
Best Local Similarity 98.9%; Pred. No. 1e-39;  
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2Y 228 GNETSVFPGTGNKTLGLAIKRFYFPKPHLPTEKFLSLQLQFDPAVIVHKQFYLTFNFEY 237  
DB 1 GNETSVFPGTGNKTLGLAIKRFYFPKPHLPTEKFLSLQLQFDPAVIVHKQFYLTFNFEY 63  
2Y 288 WFLPMKFPFIKITYEIEPLPIRNKTLSQL 316  
DB 61 WFLPMKFPFIKITYEIEPLPIRNKTLSQL 89

#### RESULT 13

AB60074 ID AB60074 standard; protein; 2703 AA.

AC AB60074;

2T 26-MAR-2002 (first entry)

2E Drosophila melanogaster polypeptide SEQ ID NO 7014.

2W Drosophila developmental biology; cell signalling; insecticide;  
2X pharmaceutical.

2S Drosophila melanogaster.

2X WO200171042-A2.

2P 27-SEP-2001.

2E 23-MAR-2001; 2001WO-US009231.

2R 23-MAR-2000; 2000US-0191637P.

2R 11-JUL-2000; 2000US-00614150.

2A (PEKE ) PE CORP NY.

2I Venter JC, Adams M, Li PWD, Myers EW;

2R WPI; 2001-656860/75.

2R N-PSDB; ABL04177.

2T New isolated nucleic acid detection reagent for detecting 1000 or more  
2T genes from Drosophila and for elucidating cell signaling and cell-cell  
2T interactions.

2S Disclosure; SEQ ID NO 7014; 21pp + Sequence Listing; English.

2C The invention relates to an isolated nucleic acid detection reagent  
2C capable of detecting 1000 or more genes from Drosophila. The invention is  
2C useful in developmental biology and in elucidating cell signaling and  
2C cell-cell interactions in higher eukaryotes for the development of  
2C insecticides, therapeutics and pharmaceutical drugs. The invention  
2C discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
2C sequences (AB101840-AB16175) and the encoded proteins (AB57737-  
2C AB572072). The sequence data for this patent did not form part of the  
2C printed specification, but was obtained in electronic format directly  
2C from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

2X Sequence 2703 AA;

Query Match 5.4%; Score 95.5; DB 4; Length 2703;  
Best Local Similarity 20.1%; Pred. No. 21;  
Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;

QY 1 RVSGIPIRHHMPVYKRPDPKPPYQAKYTPCTPGSPPIPMWGDDIVFRLQADPW 60  
DB 1966 RFNKIPORSEV-----KSKKRPHEPLWKKEKT-----PYKKGNSLRPYOLEGLNW 2032

QY 61 EFKYGDLLGHLKIMHDAIGFASSTLTGNKYTWMEYELFOLGNCPTPHLRPEMDAPWCNQG 120

DB 2033 -LKSFWYNHNCILADEWGLAKTIQSLTFHVSVEYGRG--PLVIAPLASTIPWQRE- 2088

QY 121 AACFPEGIDVHMKENGTLVQVATISGMFMQAKVQDNDETGIYETMWKASPE--- 177

DB 2089 ---FEG-----WIDMNVVYHGSVT-----SKQMQDYE--YYTTEGKYLKEPIK 2130

QY 178 -KGAETPDSYDCKSFVLRFTN-----KLAEGAEFKNTE----- 211

DB 2131 FNVLLITTFEMITVDYMDLKAENWRLCVIDEAHLKNNCKLLE-GLRQINLHREVLLSGT 2189

QY 212 ---TNYRIF--LYSGEPTVIGNETSVFGPTGNKTLGLAIKRFYFPKPHL----- 257

DB 2190 PLQNNISELPSLNFLEDSQPSQSEEFMSSEFSGLRTEBEVVKLQALLKPMMLRLKQDVE 2249

QY 258 ---PTKEFLSLQLQFDPAVIVHKQFYLTFNFEYFPMKPPFIK--ITYEIP 305

DB 2250 KSLAPKEETIEV---ELTNIQKKY-----RGILEQNSFLKKGTSANIP 2293

#### RESULT 14

ABU52308 ID ABU52308 standard; protein; 429 AA.

AC ABU52308;

2T 07-MAY-2003 (first entry)

2E Helicobacter pylori selected interacting domain (SID) protein #1651.

2X Protein-protein interaction; ulcer; selected interacting domain; SID.

2X Helicobacter pylori.

2X WO200266501-A2.

2P 29-AUG-2002.

2E 28-DEC-2001; 2001WO-EP015428.

2R 02-JAN-2001; 2001US-0259302P.

2A (HYBR-) HYBRIGENICS.

2R (INSP ) INST PASTEUR.

2I Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

2R WPI; 2002-674910/72.

2R N-PSDB; ABX70383.

2T New complexes of protein-protein interactions in Helicobacter pylori,  
2T useful for identifying modulating compounds for treating or preventing  
2T ulcers in mammals.

2S Claim 6; Page 618; 642pp; English.

2C The invention describes a complex of protein-protein interactions in  
2C Helicobacter pylori selected from 421 complexes given in the  
2C specification. The complex of protein-protein interactions are useful for  
2C screening for agents which modulate the interaction of proteins.  
2C Modulating compounds which binds to a targeted bacterial protein may be  
2C used for treating or preventing ulcers in a human or animal. This is the  
2C amino acid sequence of a selected interacting domain (SID), identified  
2C via protein-protein interactions. Note: Where the patent number printed

at the top of the pages in the specification has obscured areas of protein sequence, the indexer has replaced the residue with an X to represent an illegible residue

Sequence 429 AA;

Query Match 5.3%; Score 94; DB 5; Length 429;

Best Local Similarity 23.6%; Pred. No. 2.1; Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

```
80 FSTLTGKNTYMEWELEFOLGNCCTFPHLRPEMDAPFCN-----QGAACEFGIDVHWK 134
125 FQYTGKNSALVF-----NATTP-----WANGSIPKNSYVRFGYGVWNG 167
135 ENGLTVOYAT-----ISGNNMQAKWVKQDNETGIYETWNVKASPEKGAETWDSYDC 189
168 KTGYYTGFTADRYVITGNM-----TG-----NGAQTGGGA--T 200
190 SKFVLRTFNKLAERGAERKNIET---NYTRIFLYSGEPTTYLGNETSVPKGNKTLGLAI 246
201 INFVGATEINIA--GATFKNLKTTSQNSYMFMAIGD-----SSGSAKINVSQ 246
247 KRFY-----YFKEPHLPTEKELLSLQIFDAVIVHKQFYLF-----YNFEYV-FLPMK 293
247 SDFYDWTCGGYDFTGN-----GVFDSVNFKNAYYFQGTENSYNFKNTNFLAGN 295
294 PFPI-KITYEIRIPPIRNKTLISG 315
296 FKFGKTTIEKSVLSDASTYFPG 318
```

15

1030825

ABU0825 standard; protein; 1238 AA.

ABU0825;

19-TUN-2003 (first entry)

Protein encoded by prokaryotic essential gene #16352.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Helicobacter pylori.

MO20027183-A2.

03-OCT-2002.

21-MAR-2002; 2002MO-US009107.

21-MAR-2001; 2001US-00915242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,

NPI; 2003-029926/02.

N-PSDB; ACA34695.

New antisense nucleic acids, useful for identifying proteins or screening

for homologous nucleic acids required for cellular proliferation to

isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 58749; 1766pp; English.

of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 1238 AA;

Query Match 5.3%; Score 94; DB 6; Length 1238;

Best Local Similarity 23.6%; Pred. No. 9.7; Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

```
QY 80 FSTLTGKNTYMEWELEFOLGNCCTFPHLRPEMDAPFCN-----QGAACEFGIDVHWK 134
DB 513 FQYTGKNSALVF-----NATTP-----WANGSIPKNSYVRFGYGVWNG 555
QY 135 ENGLTVOYAT-----ISGNNMQAKWVKQDNETGIYETWNVKASPEKGAETWDSYDC 189
DB 556 KTGYYTGFTADRYVITGNM-----TG-----NGAQTGGGA--T 588
QY 190 SKFVLRTFNKLAERGAERKNIET---NYTRIFLYSGEPTTYLGNETSVPKGNKTLGLAI 246
DB 589 INFVGATEINIA--GATFKNLKTTSQNSYMFMAIGD-----SSGSAKINVSQ 634
QY 247 KRFY-----YFKEPHLPTEKELLSLQIFDAVIVHKQFYLF-----YNFEYV-FLPMK 293
DB 635 SDFYDWTCGGYDFTGN-----GVFDSVNFKNAYYFQGTENSYNFKNTNFLAGN 683
QY 294 PFPI-KITYEIRIPPIRNKTLISG 315
DB 684 FKFGKTTIEKSVLSDASTYFPG 706
```

Search completed: June 8, 2004, 14:03:58  
Job time : 84.8278 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

l protein - protein search, using sw model

on: June 8, 2004, 14:00:33 ; Search time 24.5778 Seconds

(without alignments)  
663.762 Million cell updates/sec

US-10-010-050A-2\_COPY\_31\_346

1758  
1 RVSGIPSRHWPVYKRFDF.....IKITYEILPIRMKTLISGL 316

BL0SUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

minimum DB seq length: 0

maximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/2/iaa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	91	5.2	4563	4	US-09-108-006C-1	Sequence 1, Appli
2	88.5	5.0	1462	3	US-07-792-600-31	Sequence 31, Appl
3	88.5	5.0	1462	3	US-09-157-021-31	Sequence 31, Appl
4	88.5	5.0	1462	3	US-09-156-842-31	Sequence 31, Appl
5	88.5	5.0	1462	4	US-09-591-514-31	Sequence 31, Appl
6	88	5.0	608	4	US-09-489-039A-13503	Sequence 13503, A
7	87.5	5.0	178	4	US-09-134-000C-3624	Sequence 3624, Ap
8	83	4.7	535	4	US-09-134-078-24	Sequence 24, Appl
9	83	4.7	1056	4	US-09-079-030-217	Sequence 217, App
10	83	4.7	4536	4	US-09-180-422B-27	Sequence 27, Appl
11	83	4.7	4536	4	US-09-079-030-1	Sequence 1, Appli
12	82	4.7	876	4	US-09-254-352B-19	Sequence 19, Appl
13	81	4.6	741	4	US-09-488-039A-7854	Sequence 7854, Ap
14	80.5	4.6	718	4	US-09-328-352-4640	Sequence 4640, Ap
15	80	4.6	451	4	US-09-489-039A-11849	Sequence 11849, A
16	80	4.6	607	4	US-09-537-682-1	Sequence 1, Appli
17	80	4.6	652	4	US-09-489-039A-12461	Sequence 12461, A
18	79.5	4.5	197	4	US-09-489-039A-8317	Sequence 8317, Ap
19	78.5	4.5	634	4	US-09-849-334-4	Sequence 4, Appli
20	78.5	4.5	634	4	US-10-274-878-4	Sequence 4, Appli
21	78.5	4.5	669	4	US-09-134-000C-6185	Sequence 6185, Ap
22	77	4.4	238	3	US-08-634-475-6	Sequence 6, Appli
23	77	4.4	238	4	US-09-709-791-6	Sequence 6, Appli
24	77	4.4	405	4	US-09-134-001C-5259	Sequence 5259, Ap
25	77	4.4	501	2	US-08-781-802-4	Sequence 4, Appli
26	77	4.4	501	2	US-08-781-802-6	Sequence 6, Appli
27	77	4.4	501	3	US-08-694-078-4	Sequence 4, Appli

28	77	4.4	501	3	US-08-694-078-6	Sequence 6, Appli
29	77	4.4	501	3	US-09-058-260-4	Sequence 4, Appli
30	77	4.4	501	3	US-09-058-260-6	Sequence 6, Appli
31	77	4.4	501	3	US-09-058-260-14	Sequence 14, Appli
32	77	4.4	501	3	US-09-058-260-20	Sequence 20, Appli
33	77	4.4	501	3	US-09-058-260-22	Sequence 22, Appli
34	77	4.4	501	3	US-09-058-260-24	Sequence 24, Appli
35	77	4.4	501	3	US-09-058-260-32	Sequence 32, Appli
36	77	4.4	615	4	US-09-107-532A-6507	Sequence 6507, Ap
37	77	4.4	615	4	US-09-107-532A-4897	Sequence 4897, Ap
38	76	4.3	676	4	US-09-107-532A-4897	Sequence 5806, Ap
39	76	4.3	771	4	US-09-462-284-2	Sequence 2, Appli
40	75.5	4.3	537	4	US-09-489-039A-14149	Sequence 14149, A
41	75.5	4.3	707	1	US-08-528-122-18	Sequence 18, Appl
42	75.5	4.3	707	5	PCT-US95-11720-18	Sequence 18, Appl
43	75.5	4.3	750	1	US-08-325-553-2	Sequence 2, Appli
44	75.5	4.3	750	2	US-08-394-152A-2	Sequence 2, Appli
45	75.5	4.3	750	3	US-09-044-668-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-108-006C-1  
Sequence 1, Application US/09108006C

Patent No. 6524613

GENERAL INFORMATION:

APPLICANT: Steer, Clifford J.

Kren, Betsy T.

Bandyopadhyay, Paramita

Roy-Chowdhury, Jayanta

TITLE OF INVENTION: Hepatocellular Chimeraplasty

NUMBER OF SEQUENCES: 62

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kimeragen, Inc.

STREET: 300 Pleasant Run

CITY: Newtown

STATE: PA

COUNTRY: USA

ZIP: 18940

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/108,006C

FILING DATE: 30-Jun-1992

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/054,288

FILING DATE: 30-APR-1997

APPLICATION NUMBER: 60/054,837

FILING DATE: 05-AUG-1997

APPLICATION NUMBER: 60/064,996

FILING DATE: 10-NOV-1997

APPLICATION NUMBER: 60/074,497

FILING DATE: 12-FEB-1998

APPLICATION NUMBER: PCT US 98/08834

FILING DATE: 30-APR-1998

ATTORNEY/AGENT INFORMATION:

NAME: Friebe, Thomas

REGISTRATION NUMBER: 29258

REFERENCE/DOCKET NUMBER: 7991-015-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-504-4444

TELEFAX: 215-504-4545

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4563 amino acids

TYPE: amino acid





SEQ ID NO 31

388 DAEDAMNYRGFTFPWGLANTDISYDPQKIDAQTCMAMMDNYRAGLSHQQC



NAME: McMillian, Nabeeia R.  
REGISTRATION NUMBER: P-43,363  
REFERENCE/DOCKET NUMBER: ARAG:003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 217:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1056 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-079-030-217

Query Match 4.7%; Score 83; DB 4; Length 1056;  
Best Local Similarity 19.7%; Pred. No. 5.6; Mismatches 123; Indels 66; Gaps 1;

Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 1;

```
44 MEGDDIEVFLQAPVWEFKYDGLG--HLKIMDAIGFRLTGNKNTWMEYELFQLG 100
|||
506 MDEDDFS-----KMFYISPOSSPDKKLTIFKTELKRVESDETOIKWMEEAASG 558
|||
101 NCTFPHLRPMDAPFWCQGAACFEFGIDVHWKNG-TLVQVATISGNMFQMAKWKYQ 159
|||
559 LIT--SLKDNVP-----KATGVLYDVYKVMHEHTGLTLREVSSKLRNLQNNAEWYQ 610
|||
160 DNEFTGIYETWNVKASPEKGAETWDSYDCKSFVLRTFNKLAEFGAEPKNIETN-YTRIF 218
|||
611 ---CAIRQIDIDIVRFQKASGTTGYQ-----EWDKXQNLXYQELL 649
|||
219 LYSGETVLYGNETSYPGPTGNKTLGLAIK-----RFYYPFKPHLPTKEFL 263
|||
650 TOEGASFOGLKDNVFDGLVRYTQKFMKVKHLIDSLIDFLNFPFQPGKRGITYREEL 709
|||
264 LSLQIFDAVIVHKQFY-----LFYNEFYWFLPMKFPPIK 298
|||
710 CTMF-IREVGVTVLSQVSKVHNGSILFSYFQDLVITLPEFLRK 752
|||
```

ISULT 10

1-09-180-422B-27  
Sequence 27, Application US/09180422B  
Patent No. 6444644

GENERAL INFORMATION:

APPLICANT: BRUCKDORFER, KARL R

ETTELAE, CAMILLE

TITLE OF INVENTION: ANTICOGULANT PEPTIDE FRAGMENTS DERIVED  
FROM APOLIPOPROTEIN B-100

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE, P.C.

STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/180,422B

FILING DATE: 07-Dec-1998

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36663

REFERENCE/DOCKET NUMBER: 117-268

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164000

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 4536 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 27:

US-09-180-422B-27

Query Match 4.7%; Score 83; DB 4; Length 4536;  
Best Local Similarity 19.7%; Pred. No. 54;  
Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

```
44 MEGDDIEVFLQAPVWEFKYDGLG--HLKIMDAIGFRLTGNKNTWMEYELFQLG 100
|||
396 MDEDDFS-----KMFYISPOSSPDKKLTIFKTELKRVESDETOIKWMEEAASG 4038
|||
101 NCTFPHLRPMDAPFWCQGAACFEFGIDVHWKNG-TLVQVATISGNMFQMAKWKYQ 159
|||
4039 LIT--SLKDNVP-----KATGVLYDVYKVMHEHTGLTLREVSSKLRNLQNNAEWYQ 4090
|||
160 DNEFTGIYETWNVKASPEKGAETWDSYDCKSFVLRTFNKLAEFGAEPKNIETN-YTRIF 218
|||
4091 ---CAIRQIDIDIVRFQKASGTTGYQ-----EWDKXQNLXYQELL 4129
|||
219 LYSGETVLYGNETSYPGPTGNKTLGLAIK-----RFYYPFKPHLPTKEFL 263
|||
4130 TOEGASFOGLKDNVFDGLVRYTQKFMKVKHLIDSLIDFLNFPFQPGKRGITYREEL 4189
|||
264 LSLQIFDAVIVHKQFY-----LFYNEFYWFLPMKFPPIK 298
|||
4190 CTMF-IREVGVTVLSQVSKVHNGSILFSYFQDLVITLPEFLRK 4232
|||
```

RESULT 11  
US-09-079-030-1  
Sequence 1, Application US/09079030  
Patent No. 6635623

GENERAL INFORMATION:

APPLICANT: Quevera, Jr., Juan G.

APPLICANT: Hoogseveen, Ron C.

APPLICANT: Moore, Paul J.

TITLE OF INVENTION: LIPOPROTEINS AS NUCLEIC ACID DELIVERY

TITLE OF INVENTION: VECTORS FOR TRANSFECTION OF EUKARYOTIC CELLS

NUMBER OF SEQUENCES: 229

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,030

FILING DATE: Concurrently Herewith

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: McMillian, Nabeeia R.

REGISTRATION NUMBER: P-43,363

REFERENCE/DOCKET NUMBER: ARAG:003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4536 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear  
S-09-079-030-1

Query Match 4.7%; Score 83; DB 4; Length 4536;

Best Local Similarity 19.7%; Pred. No. 54; Matches 56; Conservative 39; Mismatches 123; Indels 66; Gaps 11;

Y 44 MEGDDDEVERLQAPWMEFKYGDLLG---HLKIMHDAIGFRSTLTKNYMWEYELFQLG 100  
b 3986 MDEDDDS-----KNFYYSPSSDPDKLTTFKTELRRRESDEBQIVNNEEELASG 4038  
Y 101 NCTPPLRPMDAPFWCNGAACEFEGIDVHWKENG-TLVQVATISGNFNQMAKWKQ 159  
b 4039 LLT--SLKDNVP-----KATGVLYDVYKNKHWEHTGLTRVSSSKRRRLQNNAEWVYQ 4090  
Y 160 DNEGIYETWNNYASPEKAEETWFDSDSKVLRTRFNKLAEFGAEFKIETN-VTRIF 218  
b 4091 ----GAIRQIDDDIDVRFOKASGTTGYQ-----EMKQKQNLVQEL 4129  
Y 219 LYSGEPTYLGNETSVEFGPTGNKTLGLAIK-----REYYFPKPHLPYKEFL 263  
b 4130 TQGGQASFOGLKONVFGVLRVQKFKMKVXHLIDSLIDFLNFRPQFPQKPGVYTRREL 4189  
Y 264 LSLIQIFDAVIYHKKFY-----LFYNYFWFLPMPKFPFK 298  
b 4190 CTWF-IREVGTVLQVYSKVHNGSEILFSYFODLVITLPELBR 4232

RESULT 12  
S-09-254-352B-19

Sequence 19, Application US/09254352B  
Patent No. 6365350

GENERAL INFORMATION:  
APPLICANT: HAYASHIZAKI, Yoshihide

TITLE OF INVENTION: METHOD OF DNA SEQUENCING  
FILE REFERENCE: 024705-080  
CURRENT APPLICATION NUMBER: US/09/254,352B  
CURRENT FILING DATE: 1999-11-10  
PRIOR APPLICATION NUMBER: PCT/JP98/03039  
PRIOR FILING DATE: 1998-07-06  
PRIOR APPLICATION NUMBER: JP 10-155847  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: JP 9-196478  
PRIOR FILING DATE: 1997-07-07  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 876

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T7 RNA polymerase

S-09-254-352B-19

Query Match 4.7%; Score 82; DB 4; Length 876;

Best Local Similarity 22.3%; Pred. No. 5.5; Matches 45; Conservative 22; Mismatches 67; Indels 68; Gaps 12;

Y 35 CPTGSPFPWEGD-----DDIEVRLQAPWMEFKYDGLGLKIM----- 74  
b 356 CPVGD-VPAERSELPRPDIDITNEVARKARKREAAVVRKKAROSRRRCCEFWAQA 414  
Y 75 ----HDAIGRSTLTKNYMWEYELFQLG-----CTPPLRP-EMDAPWNC 118  
b 415 NFNANKRIWF-----PYNDWRGRYVAVSMGNDVTKSLTLAKGKYGIDGFWLWK 467  
Y 119 -QGAACFEFGIDVHWKENGTLVQVATISGNFNQMAK-----WYQDNETGIYYE 168  
b 468 IHGANC-AQVDKVPPE-----RIKFIENEGNILLASADPLNNTWMTQGDSPPAFCFE 520  
Y 169 TNNYASPEKAEETWFDSDSK 190  
b 521 YAGVK-----HGGINSYNS 535

RESULT 13

US-09-489-039A-7854

Sequence 7854, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 7854

LENGTH: 741

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7854

Query Match 4.6%; Score 81; DB 4; Length 741;

Best Local Similarity 22.7%; Pred. No. 5.5; Matches 56; Conservative 31; Mismatches 66; Indels 94; Gaps 17;

Y 63 KYGDL-GHLKIMHDAIGFRSTLTG---KNYTWMEYELFQLGNCPTPPLRPMDAPF--- 115  
b 518 KYGAANGHLPLISSVNDSDPLQIKNSYELG-----RYTDVLRTOGLAAYSTIS 570  
Y 116 -----WQNGAACFEFGIDVHWKENGTLVQVATISGNFNQMA 153  
b 571 DKTIVNRDMTIDVQSDKQRIYCEGAVDYF--IDPSW-----SVGGN-FNVL 617  
Y 154 AKWYKQDNETGIYEFWNNYK-ASPEKGAETW-----FDSYCSKFWLR 195  
b 618 KSGVQYDGR-----WQKMDVTLASPSK-ATAWGMADPWSLRVQSOQVPLSDAAGNKL 672  
Y 196 TENKLAEPGA-----EFKN-LEITNYTRIF-----LYSGEPTYLGNETSVEFGPT 238  
b 673 GNYTVDFISGYALPVGKLTFSIENLIMEDYVITMGQAPLLIS-PTY-GSSSLVBYKG 728  
Y 239 -NKTGL 244  
b 729 RGRTEGL 735

RESULT 14

US-09-328-352-4640

Sequence 4640, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 4640

LENGTH: 718

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-4640

Query Match 4.6%; Score 80.5; DB 4; Length 718;

Best Local Similarity 20.6%; Pred. No. 5.9; Matches 66; Conservative 41; Mismatches 123; Indels 91; Gaps 16;

Y 25 DPYQAKYTFPCPTSPFPWEGD--DIEVRLQAPWV-----EFYCGD 66  
b 194 DPTCYV--TYLAEALYEMGTRKDAAGLDAYKIKLGMWDIFVRNPKINGAADQENTGD 251





181	ETMEDSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	240
182	ETMEDSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	240
211	ETMFSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	270
241	TLGLAIRRFYYPPEKPHLPTEKEFLLSLQIPDAVIVHKQFYLFYVPEYWFLEPMKEPFIKIT	300
271	TLGLAIRRFYYPPEKPHLPTEKEFLLSLQIPDAVIVHKQFYLFYVPEYWFLEPMKEPFIKIT	330
301	YEEIPLPIRNKTLISGL	316
331	YEEIPLPIRNKTLISGL	346

RESULT 2

IS-10-010-050A-2

Sequence 2, Application US/10010050A

Publication No. US20020173624A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Gilbertson, Debra G.

TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

FILE REFERENCE: 97-38C1

CURRENT APPLICATION NUMBER: US/10/010, 050A

PRIOR APPLICATION NUMBER: US 09/122,363

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/053,613

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ. ID NOS: 19

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 2

LENGTH: 346

TYPE: PRT

ORGANISM: Homo sapien

IS-10-010-050A-2

Query Match

Best Local Similarity 100.0%; Score 1756; DB 13; Length 346;

Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

181	ETMEDSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	240
182	ETMEDSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	240
211	ETMFSDYDCKFVLRITGNKLAEPFAEKNTETNTRFLVSGSEPTLYGNETSVGGPTGNK	270
241	TLGLAIRRFYYPPEKPHLPTEKEFLLSLQIPDAVIVHKQFYLFYVPEYWFLEPMKEPFIKIT	300
271	TLGLAIRRFYYPPEKPHLPTEKEFLLSLQIPDAVIVHKQFYLFYVPEYWFLEPMKEPFIKIT	330
301	YEEIPLPIRNKTLISGL	316
331	YEEIPLPIRNKTLISGL	346

RESULT 3

IS-10-653-595-126

Sequence 126, Application US/10653595

Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

```

FILE REFERENCE: P2027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIORITY APPLICATION NUMBER: US 09/397945
PRIORITY FILING DATE: 1999-09-17
PRIORITY APPLICATION NUMBER: PCT/US99/05804
PRIORITY FILING DATE: 1999-03-18
PRIORITY APPLICATION NUMBER: 60/078,566
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,576
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,573
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,574
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,579
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/080,314
PRIORITY FILING DATE: 1998-04-01
PRIORITY APPLICATION NUMBER: 60/080,312
PRIORITY FILING DATE: 1998-04-01
PRIORITY APPLICATION NUMBER: 60/078,578
PRIORITY FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
Seq ID NO 126
LENGTH: 346
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (242)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (246)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-653-595-126

```

Query Match	Best Local Similarity	99.4%	Score 1747	DB 12	Length 346
Matches 314	Conservative 0	Mismatches 2	Indels 0	Gaps 0	
QY	1	RVSGIPSRHRMPVYKREDFPRKEDPYCOAKYTCFPTSSPIFVMEGGDDIEVERLQAPVW	60		
Db	31	RVSGIPSRHRMPVYKREDFPRKEDPYCOAKYTCFPTSSPIFVMEGGDDIEVERLQAPVW	90		
QY	61	EFKYGDLLGHKIKMHDAIGFRSTLTGKYTMEWYELPOLGNTFPHLRPEMMAFPCNCG	120		
Db	91	EFKYGDLLGHKIKMHDAIGFRSTLTGKYTMEWYELPOLGNTFPHLRPEMMAFPCNCG	150		
QY	121	AACFEFGIDDVHMKENGTLVQVATISGNMFQMAKVVQDNDETGIYYETVNVKASPEKGA	180		
Db	151	AACFEFGIDDVHMKENGTLVQVATISGNMFQMAKVVQDNDETGIYYETVNVKASPEKGA	210		
QY	181	ETWEDSYDSCSFVLTFRPKLAEPGAEFKNIEFNTRYIRILYSEFPYTLGNETSVFGTGAK	240		
Db	211	ETWEDSYDSCSFVLTFRPKLAEPGAEFKNIEFNTRYIRILYSEFPYTLGNETSVFGTGAK	270		
QY	241	TLGAIKRFYFPRFKPHLPTKEFLLSLQIDPAVIVHKOPLYLNFNEEYWLPMKKFPPIKIT	300		
Db	271	TLGAIKRFYFPRFKPHLPTKEFLLSLQIDPAVIVHKOPLYLNFNEEYWLPMKKFPPIKIT	330		
QY	301	YEETPLPIRNTKLTGSL	316		
Db	331	YEETPLPIRNTKLTGSL	346		

## GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc. et al.  
TITLE OF INVENTION: 95 Human secreted proteins  
FILE REFERENCE: P20272P1

CURRENT FILING DATE: 1999-09-17  
CURRENT APPLICATION NUMBER: US/09/397,945

PRIOR FILING DATE: 1999-09-17  
PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18  
PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,581

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,577

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/078,563

PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: 60/080,313

PRIOR FILING DATE: 1998-04-01  
NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 126

LENGTH: 347

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (242)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (246)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (347)

OTHER INFORMATION: Xaa equals stop translation

3-09-397-945-126

Query Match 99.4%; Score 1747; DB 12; Length 347;  
Best Local Similarity 99.4%; Pred. No. 6.3e-166;

Matches 314; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
1 RVSGIPSRHMPVVKRPDPKPPYQQAQYTCPTGSPSPVWEGDDIDVFLQAPVW 60
  |||
31 RVSGIPSRHMPVVKRPDPKPPYQQAQYTCPTGSPSPVWEGDDIDVFLQAPVW 90
  |||
61 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFQNGCTEPHLPENDAFW 120
  |||
91 EFKYGDLLGHLKIMHDAIGFRSTLTGKNYTWEMWELFQNGCTEPHLPENDAFW 150
  |||
121 AACFPEGIDDVHAKENGTLVAVATISGNMFQMAKWKQODNETGIYYETWVKSPE 180
  |||
151 AACFPEGIDDVHAKENGTLVAVATISGNMFQMAKWKQODNETGIYYETWVKSPE 210
  |||
181 EFWEDSDYDCSFVLRTEFKLAEPGAEPFNIEFTNTRLEIYSGEPYVIGNETSV 240
  |||
211 EFWEDSDYDCSFVLRTEFKLAEPGAEPFNIEFTNTRLEIYSGEPYVIGNETSV 270
  |||
241 TLGLAIKRFYVPKPHLPTEKEFLSLQIPDAVTHKQFYLFINFEWFLPMKEP 300
  |||
271 TLGLAIKRFYVPKPHLPTEKEFLSLQIPDAVTHKQFYLFINFEWFLPMKEP 330
  |||
```

OY 301 YEELPLIRNKTLISGL 316  
Db 331 YEELPLIRNKTLISGL 346

## RESULT 5

US-09-864-761-39378

Sequence 39378, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecolica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1

SEQ ID NO 39378

LENGTH: 163

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE: MAP TO AC00126.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3

OTHER INFORMATION: EST\_HUMAN HIT: H01255.1, EVALUATE 5.00e-59



```
; OTHER INFORMATION: SWISSPROT HIT: 075503, EVALU 2.00e-96
JS-09-864-761-39378
Query Match      50.4%; Score 886; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 2.7e-80;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 154 AKWVKQDNENGIYETNNVVASPEKGAETWEDSDYDCKFYLTETNKLAERGAFFKNIETN 213
    1 AKWVKQDNENGIYETNNVVASPEKGAETWEDSDYDCKFYLTETNKLAERGAFFKNIETN 60

2y 214 YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAY 273
    61 YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIKRFYPPKPHLPTEKFLSLIQIFDAY 120

2y 274 IVHKQFLYFNFEYWFLEPMKPPFIKITYEELPIPIRNKTLISGL 316
    121 IVHKQFLYFNFEYWFLEPMKPPFIKITYEELPIPIRNKTLISGL 163

RESULT 6
JS-10-264-049-4034
; Sequence 4034, Application US/10264049
; Publication No. US2004000579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264, 049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4034
; LENGTH: 89
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
JS-10-264-049-4034

Query Match      26.8%; Score 472; DB 15; Length 89;
Best Local Similarity 98.9%; Pred. No. 2.9e-39;
Matches 88; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2y 228 GNETSVGPGNCTGLAIKRFYPPKPHLPTEKFLSLIQIFDAYIVHKQFLYFNFEY 287
    1 GNETSVGPGNCTGLAIKRFYPPKPHLPTEKFLSLIQIFDAYIVHKQFLYFNFEY 60

2y 288 WFLPMKPPFIKITYEELPIPIRNKTLISGL 316
    61 WFLPMKPPFIKITYEELPIPIRNKTLISGL 89

RESULT 7
JS-10-369-493-12714
; Sequence 12714, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
```

```
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12714
; LENGTH: 328
; TYPE: PR1
; ORGANISM: Aspergillus nidulans
US-10-369-493-12714

Query Match      5.5%; Score 97.5; DB 15; Length 328;
Best Local Similarity 26.1%; Pred. No. 0.47;
Matches 41; Conservative 19; Mismatches 50; Indels 47; Gaps 7;

2y 138 TIVQVATISGNMFMQMA-----KMWKQDNENGIYETNNVVASPEKGAET 182
    57 TIVSIGFQSNHTRQVAAVARKGLEVALVQEKWDMKNGVYDAGNITQSLRMGADT 116

2y 183 WEDSYDCKRFVLRTEFKLAFFGAE---PKNIETNYTRIFLYSGEPTLYGNETSVEGPTG 238
    117 RLD-----MGFGIEHKLTLKNE---EELIRAGGRKYI-----PAGASD 154

2y 239 NKTGLAIRFYPPKPHLPTEKFLSLIQIFDAYIV 275
    155 HPLGGLGFAR--WAFEVEMQEK---LGVFFDTVIY 185

RESULT 8
US-10-408-765A-58
; Sequence 58, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale R.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 836
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-408-765A-58

Query Match      5.4%; Score 95.5; DB 16; Length 836;
Best Local Similarity 20.5%; Pred. No. 2.7;
Matches 61; Conservative 39; Mismatches 125; Indels 73; Gaps 12;

2y 37 TQSPPIV-----MEGDDIEVFRLOAPVMEKFGYDLG---HLKIMDAIGFRSTLNG 86
    272 TSAFSPAAGTGMDDNEDDDFS-----KMFYFPOSSPDKLITFTKTLRVRSDSE 324

2y 87 KNYTWWEYELFOLGNCPTPHLRPEMDAPFWCNGAACFEFGIDDVHMKENG-TLVQVAT 145
    325 TQIKVWEEEAASGLIT--SLKDNV-----KATGVLDVYVKYHMEHTGTLRVRSSK 376

2y 146 SGMEFQMAKMYKQDNENGIYETNNVVASPEKGAETWEDSDYDCKFYLTETNKLAERGA 205
    377 LRRLICDHAWEYQ---GAIREIDIDERFQKAGAGTGTQV----- 415

2y 206 EFKNIETN-YTRIFLYSGEPTLYGNETSVEGPTGNKTLGLAIK-----RF 249
    416 EMDKAKNLYQBELILQEGASFGGLKDNVFDGLVVRTOGFHMKVKKLIDSLIDFLNPPRF 475

2y 250 YPFKPHLPTEKFLSLIQIFDAYIVHKQFY-----LFYNEBYWFLPMKPPFIK 298
```



[illegible]

## RESULT 11

US-10-424-599-262855  
Sequence 262855, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21 (53223) B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 262855  
LENGTH: 347  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_7937C.1.pep  
US-10-424-599-262855

Query Match	5.28;	Score 91.5;	DB 12;	Length 347;
Best Local Similarity	23.9%;	Pred. No. 2;		
Matches	52;	Conservative	25;	Mismatches 72;
				Indels 69;
				Gaps 12;

```

2y 58 PWEKSEYKGLLGHLLKIMDAIGFSTLLGKUYMEWELEFOLGCTSPHRLREPDCAFWC 117
2b 47 PRMEG-GRALD-----RSKNKRMIWLWED-----GPCDPEDEGSRNGCLNY 90
2y 118 NCGAACFEEGIDVDHMKENGTLVVOYATISGNENOMAKMYKODNETSI---YYETMANVKA 174
2b 91 KEGADVFPEKSHSEVYSYGEGLMGKVLADNNH-----KMYVSDPTQNGCESSYIGAMN--A 142
2y 175 SEPKAEITMFDSDYDSKFV-----LRTPKLALEPAEKRIETNYTRFLY 220
2b 143 SMDHOPKTM--EFOLNSGIQTIAVIAAREGLVQGSFKI---AEDLNFWVSIGORF-- 194
2y 221 GSEPTIYIGNETSVEFPTNKNKTLGLAIKRFYFPKPHLP 258
2b 195 ----SYLHSIPGVF-----SIQR-----PHLP 212

```

RESULT 12

JS-09-374-046A-160  
 Sequence 160, Application US/09374046A  
 Publication No. US20030096951A1  
 GENERAL INFORMATION:  
 APPLICANT: Jacobs, Kenneth  
 APPLICANT: McCoy, John M.  
 APPLICANT: Lavallie, Edward R.  
 APPLICANT: Collins-Racie, Lisa A.  
 APPLICANT: Evans, Cheryl  
 APPLICANT: Merberg, David  
 APPLICANT: Treacy, Maurice  
 APPLICANT: Agostino, Michael J.  
 APPLICANT: Steininger II, Robert J.  
 APPLICANT: Spaulding, Vikki  
 APPLICANT: Wong, Gordon G.  
 APPLICANT: Clark, Hilary  
 APPLICANT: Pechtel, Kim  
 APPLICANT: Genetics Institute, Inc.  
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
 FILE REFERENCE: GI 6075-83A

```

:
:
: CURRENT APPLICATION NUMBER: US/09/374,046A
: CURRENT FILING DATE: 1999-08-13
: NUMBER OF SEQ ID NOS: 240
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 150
: LENGTH: 742
:
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-374-046A-160

```

Query Match	5.2%;	Score 91;	DB 10;	Length 742;
Best Local Similarity	22.1%;	Pred. No. 6.4;		
Matches 62;	Conservative 39;	Mismatches 105;	Indels 74;	Gaps 16;

QY	86	GKRYTMEWYLLFOJGNTSPFH-----LREBMAF-FWNCOG-----AACFEGIDVHWK	134
Db	325	GSAYQFHSMWVFLV-YCAFPSVALGALTQPPSPRFFLENGKHDEAMWMLKQVHDINMR	383
QY	135	ENGTLLVOVATISG-NMFNQAKVKKQDNETGIYETWNVKASPEKAEITFDSYDSKFLV	193
Db	384	AKGPEFVEFVTHIKTHQDEDELEIIOSTDGTWYQORGVAL-SLGQVWGNPLSC-----	438
QY	194	LRTENKLAERGEAEKKNLETN-----YRIFLYSGEPYTL-----GNETSVEF--	234
Db	439	FGPEYKRRITLMMKGWFLMSFSYIGLTYWPFDMTRHDAVLYASATKTFPG	489
QY	235	-----GPTGNKTLGIALKRPYYFKPK-----HLPTKEFLLSLQ--IFDAVI VHKQFY--	280
Db	490	ERVEHWTFNFTLHNOIHRGQGYFNDKETGLRLKSVSPEDSLFECYEPEDVTSNSTPFRNC	549
QY	281	-----LFYN-----PEYWFLLPMKFPPIKITY-----EELPLPI 308	
Db	550	TFINLYWFYNTDLFEYKTFVNSR--LINSTFLANKEGCLDV 587	

## RESULT 13

```

US-10-616-263-1b0
/ Sequence 160, Application US/10616263
/ Publication No. US20040038276A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: LaVaille, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steinger II, Robert J.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Genetics Institute, Inc.
/ TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
/ FILE REFERENCE: 00766.000103.5
CURRENT APPLICATION NUMBER: US/10/616,263
CURRENT FILING DATE: 2003-07-08
NUMBER OF SEQ ID NOS: 240
/ SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 160
LENGTH: 742
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-616-263-160

```

Query Match	5.28;	Score 91;	DB 12;	Length 742;
Best Local Similarity	22.1%;	Pred. No. 6.4;		
Matches 62;	Conservative 39;	Mismatches 105;	Indels 74;	Gaps 16;

Qy 86GKNYTMVEYELTOLGNTCFPH-----LRPDADP-FWNCNOG---AACFEFGIDDDHWK 134  
Db 325GSAYQHFHSRVRVFL-VCAFPSPYALGTLTQOESPRFLFNKDHDEAMMVLKQVHDINMR 383

```

135 ENGLVOVATISG-NMENQAKWVKODNENGIGYETWNVKASPEKGAETWFDSDCKFV 193
1384 AKGHEPESVSTHKTTHQDELEIJDSDGTWQVNGVRL-SLGGQVWGNFLSC----- 438
194 LRTNKLAEFGAEFKNIETN-----YTRIFLYSGEPYTL-----GNETSVF-- 234
439 -----FQPEYRRLTLMMGVWFTMSFSYGLTWPFDPMIRHLQAVDASRTKFPFG 489
235 -----GPTGNKTLGLAIKRFYFPEKP-----HLPTKEFLSLIQ--IPDAVIVHKQFY--- 280
490 ERVEHVTFNFTLENOIHGGQYFNDKFIGRLKSVSEDSLFEECYFEDVTSNTFFRNC 549
281 -----LFTN---FHYWFLPMKFPFIKITV---EELPLPI 308
550 TFINTVFYNTDLFEYKFNVSRL--LINSTFLNKGCCPLDV 587

;SUIT 14
;-10-408-765A-2332
Sequence 2332, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Faly, Boia D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary W.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408.765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2332
LENGTH: 742
TYPE: PRT
ORGANISM: Homo sapiens
;-10-408-765A-2332

Query Match 5.2%; Score 91; DB 16; Length 742;
Best Local Similarity 22.1%; Pred. No. 6.4; Indels 74; Gaps 16;
Matches 62; Conservative 39; Mismatches 105;

86 GKNTYMEVYELFQNGCTFPH-----LRPMDAP-FWCNOG---AACFPEGIDVHWK 134
325 GSATQPSHWRVFL-VCAFPSEVFAIGALTTPESPFRFLENGKHDEAMVILKQVHDNMR 383
135 ENGLVOVATISG-NMENQAKWVKODNENGIGYETWNVKASPEKGAETWFDSDCKFV 193
384 AKGHEPESVSTHKTTHQDELEIJDSDGTWQVNGVRL-SLGGQVWGNFLSC----- 438
194 LRTNKLAEFGAEFKNIETN-----YTRIFLYSGEPYTL-----GNETSVF-- 234
439 -----FQPEYRRLTLMMGVWFTMSFSYGLTWPFDPMIRHLQAVDASRTKFPFG 489
235 -----GPTGNKTLGLAIKRFYFPEKP-----HLPTKEFLSLIQ--IPDAVIVHKQFY--- 280
490 ERVEHVTFNFTLENOIHGGQYFNDKFIGRLKSVSEDSLFEECYFEDVTSNTFFRNC 549
281 -----LFTN---FHYWFLPMKFPFIKITV---EELPLPI 308
550 TFINTVFYNTDLFEYKFNVSRL--LINSTFLNKGCCPLDV 587

```

```

; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 128
; LENGTH: 4563
; TYPE: PRT
; ORGANISM: Homo sapiens
;-09-870-759-128

Query Match 5.2%; Score 91; DB 9; Length 4563;
Best Local Similarity 20.4%; Pred. No. 75; Indels 122; Gaps 11;
Matches 58; Conservative 38; Mismatches 122;

QY 44 MEGDDIEVFRLQAPVWBFKYGDLG---HLKIMHDAIGRSTLTGKNTMEVYELFQLG 100
DB 4013 MDEDDDFS-----KMFYYSFQSSPDKLTITFKTLRRESDEPQIKNWEBAASG 4065
QY 101 NCTFPHLRPMDAPFWCNOGAACFEESIDDVHWKENG-TLVQVATISGNNENQAKWVKQ 159
DB 4066 ILTF-SLKNVVP-----KATGVLYDVYNNKXHWHTGLTLREVSSKLRRLQDHAEMVYQ 4117
QY 160 DNENGIYETWNVKASPEKGAETWFDSDCKFVTLRTFNKLAEFGAEFKNIETN-YTRIF 218
DB 4118 ---GATREIDDDIDERPOKASGTTGTQ-----EWDKAQNTLYOELL 4156
QY 219 LYSGEPTYLGNETSVPFGPTGNKTLGLAIK-----RFYFPEKHLPTKEFL 263
DB 4157 TQEGQASFOGLKNDVFEGLYRVTOEFEMKYKHLIDSLIDPLNFRFQPGKRGITREBEL 4216
QY 264 LSLQLPDAVIVHKQFY-----LFYNEFYWFLPMKFPFIK 298
DB 4217 CTMP-IREVGTVLSQVYSKVHNGSEILFSYFQDLVITLPELRK 4259

Search completed: June 8, 2004, 14:11:58
Job time : 65.9556 secs

```

```

;SUIT 15
;-09-870-759-128
Sequence 128, Application US/09870759
Patent No. US2002017551A1
GENERAL INFORMATION:

```



130 SKVARIENLAE GAERKNIET---NIRKFLIDSEFTIIONEISVEGFIGNNUSDAI 218

A:Map position: 5  
A<sub>1</sub>introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1335/1.

Query Match 5.3%; Score 93; DB 2; Length 3036;  
Best Local Similarity 23.2%; Pred. No. 44; Mismatches 76; Conservative 40; Indels 102; Gaps 2;

Y 72 KIMHDAIGRSTLTGKNYTMEWELFQLGNCPTP-----HLRPENDADFWCNOGAACEF 125  
||| : : :  
1574 KILLGGVAFAMISNNFTLN-----TAPKPGWEIR-GNEKNLICRRNFFRN 1622

Y 126 EGIDDVHM--KENGT---LVQVATTSGNMENQAKVKODNETGIYTEFNWKASPEKGA 180  
||| : : :  
1623 WG----HWMIKLDATSQYLRLQIDVPSPYOYNLEYE-NRFNRQGDVDWM-----PRSYA 1672

b 181 EFMEDPSYDSCSKVLRTFKNLAEF---GAEFKI-ETNT-----RIFLX 220  
||| : : :  
1673 LCVFGSQQKEVAHFHNREFNELIFELIVSGAKYSDFESTMTHTNHMGTCGEAVIAZGVFPF 1732

b 221 SG-----EPITYGNETSU-----EGDTG-----KITGLA 245  
||| : : :  
1733 DDMNTPTRAEMLPFYVSNDLSINFMWNPMRDGLANAYLIEPVHDLAGRVIEDKNTLLI 1792

y 246 IKRF-----YYFEKPHELPTEKFEL---SLIQTFDAVIH---KOFYLEYNF-----EW 288  
||| : : :  
1793 TERWTSFYNYNPFRERYRITRODTIMPGRATLYLEQVVEHVMPNVAILVLGNLVARGEV 1852

Y 289 FLPMKEPPFIKT---YE-EIPILPIRK 311  
||| : : :  
b 1853 -QPIRKPINVTVMQKGSEIPTERYKK 1879

ESUT 6  
70316  
ature 392, 353-358, 1998  
onserved hypothetical protein aq\_175 - Aquifex aeolicus  
Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 31-Mar-2000  
Accession: F70316  
Decort, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
  
Reference number: A70300; NCBI:9819666; PMID:9537320  
Accession: F70316  
Status: preliminary; nucleic acid sequence not shown; translation not shown  
Molecule type: DNA  
Residues: 1-562 <AO>  
Cross-references: GB:AEO00677; NID:g2982900; PIDN:AAC06534.1; PID:g2982916; GB:AEO0065  
Experimental source: strain VFS  
Genetics:  
Gene: aq\_175  
Superfamily: probable 60k inner membrane protein; stage III sporulation protein homolo  
;293-486/Domains: stage III sporulation protein homology <SPOR>

Query Match 5.2%; Score 92; DB 2; Length 502;  
Best Local Similarity 21.3%; Pred. No. 5.4; Mismatches 63; Conservative 51; Indels 114; Gaps 15;

Y 40 PLPWEGDDDIIEVFRLQAFAVEREKYGDLGHKLKIMHDAIGRSTLTGKNYTMEWELFOL 99  
||| : : :  
118 PLEIFNGNDLDQ-KINPGEYELIKEG--NSVELIHKEKLVKTLISYKNGAL----- 166

Y 100 GNCTPPLHPF-MDAFWCNQGA---ACFFBGIDIVHKENGTVLVQVATISGMNERQMA 154  
||| : : :  
167 -----HLTVEGILKPFWFVGSPDPDEAFYTHVGVB-LKINEVVALLDY----- 203

Y 155 KWVKONENGIYETTNVNAASPDKAGETWMDSDYDSGFVLRRTNKLAEPGAEFNITEY 214  
||| : : :  
210 ----DDLKGINEFBGNIPEGGESRYFFRGAAQVOQHAIYK---VKDGDKVSVLSST- 258

Y 215 TRIFYSGEPT-YLG-----NETSVFGTGTKTGLALK--REYYPKPKPHLPTKEFL 263  
||| : : :  
b 259 ---FLTDGKTIYLGAKDVARURELGLAVPLIDWGTLIKIYKPLFLDWIYEH--TGSW 313

Y 264 LSLL-----QLFDANIYHQGYLFNFEEYWFLPMKEPPFIKTILEBELPDPIRKTL 313  
||| : : :

DB 314 LSIILVLTIVIRFLPPLIGKSVSMQKLQE--LAPRMEIKKKYKXDDPVAKGEMM 367

RESULT 7  
T21702  
Hypotheoretical protein F33E2.5 - Caenorhabditis elegans  
C.Species: Caenorhabditis elegans  
C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C.Accession: T21702  
R.Lennard, N.  
submitted to the EMBL Data Library, January 1997  
A.Reference number: Z19461  
A.Accession: T21702  
A>Status: preliminary; translated from GB/EMBL/DDBT  
A.Molecule type: DNA  
A.Residues: 1-291 <WILD>  
A.Cross-References: EMBL:Z84574; PIDN:CAB06543.1; GSPDB:GN00019; CESP:F33E2.5  
A.Experimental source: clone F33E2  
C.Genetics:  
A.Gene: CESP:F33E2.5  
A.Map position: 1  
A.Introns: 252/1

Query Match            5.2%; Score 91; DB 2; Length 291;  
Best Local Similarity 17.9%; Pred.No.3.3;  
Matches 57; Conservative 47; Mismatches 92; Indels 122; Gaps 13;

QY 27 YCAQKYTECP---TGSPPIVMEGDIDIEVFRL-QAPVMEFYKGDLGLHKIMDAIGFR 81  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 25 YEARYPTPIKLIKKTADPQKLEGSKDRIATFTLSGEPIADKCKTEMKNVLDEQ----FQ 80  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 82 STLTKNY-----TWWEYELFOLNGCTPEPHLRPEMDAPFCNOGA 121  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 81 DDIIPANIYAVALMA.PNFKIGQNCHEFDPLDWYNKIK-GYFTYKKITEQVAAYFNDEAG 139  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 122 ACFFEGIDDVHKENGTLVQAATISGMFMNQAAKNYK--ODNEGITYETMNVAASPEKG 179  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 140 NQFLGV-----CYATPAG-----KWAKQLQTEOTLLRFEE--QKAEP-- 176  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 180 AETMFSDYSCFEVLRTFNKLAEFGAEPKNIETNRYRIFLYSGEPTLVGNETSVEFGPTGN 239  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 177 ---FPLNKSKEPDTKAFKKPARYG--RCETEGCRVLVAVGKA----- 214  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 240 KTLGLAIRKFYYPFKPHLPTEKELLSLQTFDAVIHKQPYLFY-----NF 285  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 215 -----DQIYHLNVHPVYCMQWTRGCACACAKDH 243  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 286 EYWFLPMKEPFIKIYYEE 303  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 244 EYMWIAPEEPVVOKSKEE 261  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 8  
DB5056  
probable aspartic proteinase [imported] - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
C.Accession: DB5056  
R.anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Harbor Laboratory  
Nature 402, 769-777, 1999  
Article Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A.Reference number: AB5001; MIDID:20083488; PMID:10617198  
A.Accession: DB5056  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-508 <STO->  
A.Cross-References: GB:INC\_001268; NID:g7267203; PIDN:CA577914.1; GSPDB:GN00140  
C.Genetics:  
A.Gene: ATAG04460  
A.Map position: 4  
C.Superfamily: oryzasin; saposin repeat homology

Query Match            5.2%; Score 91; DB 2; Length 508;

Best Local Similarity 23.8%; Pred. No. 6.7;  
Matches 36; Conservative 17; Mismatches 48; Indels 50; Gaps 7;  
27 YCOAKY-----FPCPTGSPDPVMEG-----DDDIIVFRLQAPVMEF-----62  
125 YFHSKYASOSSSSRRKNGKPAISIRYGTGALSISFNSNDVAVGDIIVKEQFIETSPGI 164  
2Y 63 -----KXGDDLGLKIMHDAIGFRSTLTGKNYTWMEYLEFOLGNCPTPHLRPEMDAPFC 117  
185 TPLAKNGDILG-----LGFKEISVG-NSFVWVWVWVEKG-----LVKEPISFNL 229  
2Y 118 N-----QGAACFPEGIDVHKMGNTLVQY 142  
230 NRNPDKPEGGEIVFGVDPKHFKEHTFV 260

RESULT 9  
TC1486  
neopullulanase (EC 3.2.1.135) - Thermoactinomyces vulgare  
A:Alternate names: alpha-amylase II  
C:Species: Thermoactinomyces vulgare  
C>Date: 31-Dec-1993 #sequence\_revision 18-Aug-1995 #text\_change 16-Jun-2000  
C:Accession: JCI1486  
C:Ionozuka, T.; Ohtsuka, M.; Mogi, S.; Sakai, H.; Ohta, T.; Sakano, Y.  
3:Biochem. Biotechnol. 57, 395-401, 1993  
A:Title: A neopullulanase-type alpha-amylase gene from Thermoactinomyces vulgare R-47.  
A:Reference number: JCI1486; PMID:93222535; PMID:7763540  
A:Accession: JCI1486  
A:Molecule type: DNA  
A:Residues: 1-585 <TON>  
A:Cross-references: GB:D13178; NID:g391625; PIDN:BAA02473.1; PID:g398125  
A:Function:  
A:Description: hydrolysis of alpha-(1->4)-glucosidic linkages of pullulan to produce pan  
A:Pathway: pullulan degradation  
A:Note: also has alpha-amylase activity  
C:Superfamily: neopullulanase; alpha-amylase core homology  
C:Keywords: glucosidase; hydrolase; polysaccharide degradation  
C:293-424/Domain: alpha-amylase core homology <AMY>  
C:329,354,421/Active site: Glu, Glu, Asp #status predicted

Query Match 5.1%; Score 90; DB 1; Length 585;  
Best Local Similarity 21.5%; Pred. No. 9.7;  
Matches 53; Conservative 30; Mismatches 68; Indels 96; Gaps 15;

2Y 71 LKIMHDAI-----GFRSTLT--GKNYTMW-----YELFOLGNCFT 104  
234 IKLIDAVFPHAGQFAFPFADVLQKGQSRKDMFTEDPVSQTSKTNVETRAVQVPM 253  
2Y 105 PHLR---PEMD-----APFWCNOGAACFPEG-----IDVHKMGNTLVQV---A 143  
234 PKLRTNPEVKEVLEFDVARFMEQ-----IDGWRIDVANEVDHAFMRFRRLVXSLNPDA 349  
2Y 144 TISGMMNCAKAVKQKONNGI-----YETFWNVKAS-----PEKG 179  
350 LIWEIWIHDSAGWIMGQFDSVWVNYLFRESYIRFAGELHAERPDALTRARMLYDEQA 409  
2Y 180 AE-TW--FDSYDCKFVLRFTFNKLAEGAEFKNIETVTRI-----FLYSGEP--TYLGNE 230  
234 AOGIMNLLDSDHTRFLITSCGNGNAKF-----RLAVLFQMTYIGTPIIYVGDE 457  
2Y 231 TSVEGPT 237  
458 IGMAGAT 464

RESULT 10

T27404  
hypothetical protein Y75B8A.22 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27404  
C:Barlow, K.  
submitted to the EMBL Data Library, November 1998

A:Reference number: Z20361  
A:Accession: T27404  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1353 <MIL>  
A:Cross-references: EMBL:AL035514; NID:e1343251; PIDN:CAA22106.1; CESP:Y75B8A.22  
A:Experimental source: clone Y75B8A  
C:Genetics:  
A:Gene: CESP:Y75B8A.22  
A:Introns: 26/1; 79/2; 634/3; 996/1; 1134/1; 1296/3

Query Match 5.1%; Score 89; DB 2; Length 1353;  
Best Local Similarity 19.1%; Pred. No. 34;  
Matches 73; Conservative 51; Mismatches 124; Indels 134; Gaps 18;

QY 1 RVSGIP-----SRHMPVVPYK--FPRRPDPYCO-----AK 31  
DB 607 RSRGFPAVGLVHNSALMPESFKGLTFQDSPGSEDOQLLEQLKADMKVAKDYK 666  
QY 32 YTECPPTGSPDPVMEGDDIEVFLQAPVMEFKYDGLGLKIMHDAIGFRSTLTGKNYTM 91  
DB 667 AESCKTCDEDPAYKKYDKDQATLQS-LWE-QSTDTLARI-LSHELPESEST-SPVWML 722  
QY 92 EWELEFOLGNCFTPHLRPEMDAPF-----KCNQGAACF 124  
DB 723 D-----ITPVOQKFAMLAIQALRAPDLPAVGLYHTSRKLWFGDEAIFG 768  
QY 125 FEGIDVHKMGNTLVQVATISGMM--FNQAKVVK-----QDNETGIYETWNVKAS 175  
DB 769 APGI-----GVSEIEMELKALIEADLHEVARREKVAEDRADDDEBPAPPYDSEQ 820  
QY 176 PEKGAETW-----FDSYDCKFVLRFTFNKLAEGAEFKNIETVTRIIFYSGEPTYL 227  
DB 821 EEEVFPAMKVEIDFQDSYVC-----KFSNDVLKMYVFLIN---DFS 861  
QY 228 GNTSVFPGPTGNTGLAKRFRYPKPHLPTEKFLSLQITDAIVH-----KQ 278  
DB 862 KNSTEL-----NVALVKMRLAIFDK-LPIKLQVSLFQVSKVNEHTHLSKDLRKS 914  
QY 279 FYLFNFEYWFELPMKPFITKIT 300  
DB 915 SRLYEIYQGFHLKKFPFSKFT 936

RESULT 11

T51094  
acid phosphatase (EC 3.1.3.2) purple 1, precursor [similarity] - sweet potato  
C:Species: Ipomoea batatas (sweet potato)  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 22-Oct-2001  
C:Accession: T51094  
C:Durum, A.; Eicken, C.; Spener, F.; Krebs, B.  
Biochim. Biophys. Acta 1434, 202-209, 1999  
A:Title: Cloning and comparative protein modeling of two purple acid phosphatase isozymes  
A:Reference number: Z25392; PMID:20028255; PMID:10556574  
A:Accession: T51094  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-465 <DUR>  
A:Cross-references: EMBL:AU06224; PIDN:CAA06921.1  
C:Genetics:  
A:Gene: pap  
C:Superfamily: kidney bean purple acid phosphatase; phosphoesterase core homology  
C:Keywords: phosphoric monoester hydrolase

Query Match 5.0%; Score 88.5; DB 2; Length 465;  
Best Local Similarity 21.4%; Pred. No. 9.7;  
Matches 51; Conservative 31; Mismatches 77; Indels 79; Gaps 13;

QY 131 VHWKNGTIVQVATISGMM-----FNQAKVVK-----DNETGIYET-----W 170  
DB 87 VVWSENSQHKVA--KGNIRTYVFNVTSGYIHICTIRNLEVTXKYVEVIGIGNTRSF 144  
QY 171 NVKASPEKGAETWFDSDSKFVLRFTFNKLAEGAEFKNIE--TNVTR-----IFLYSG 222



b 145 -FTTPEVDPD-----VPYTFGLIGDLSGFSFDSNRTLTHERNPIKQAVLFEVG 192  
 Y 223 EPTVYGN-----ETSVGCP-----TGNTKTGLA-----IKRF 249  
 b 193 DLSYADNYNHNHNDWMTGRFVENSATQPMWTAQNEIIDPAPETIGETKPKFPTKX 252  
 Y 250 YYPFKEHLPTEKELLSLQIFDAVIVHKQF--YLFYNEFYWFLPMKFPFIKITYEIP 305  
 b 253 HVPYKASGSTETFTFWYSIKRASAAYITVLSYSYAVGKTPQYKXLEELP--KYNRTETP 308

## RESULT 12

JHUAC

NA-directed DNA polymerase (EC 2.7.7.7) alpha catalytic chain - human

Species: Homo sapiens (man)  
 Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 27-Oct-2003  
 Accession: S00257; A50440; I57513; S12665  
 Wong, S.W.; Wahl, A.F.; Yuan, P.M.; Pearson, B.E.; Arai, K.I.; Korn, D.; Hun  
 MBO J. 7, 37-47, 1988

Title: Human DNA polymerase alpha gene expression is cell proliferation dependent and  
 Reference number: S00257; MUID:88196090; PMID:3359994

Accession: S00257  
 Molecule type: mRNA

Residues: 1-1462 <WON>  
 Accession: A30440

Molecule type: protein  
 Residues: 438-449;495-502; 'G', 838-848;1090-1105;1201-1216;1397-1407;1444-1453 <MON2>  
 Pearson, B.E.; Nashener, H.P.; Wang, T.S.

Ol. Cell. Biol. 11, 2081-2095, 1991

Title: Human DNA polymerase alpha gene: sequences controlling expression in cycling an  
 Reference number: I57513; MUID:91172197; PMID:2005899

Accession: I57513  
 Status: translation not shown; translated from GB/EMBL/DBJ

Molecule type: DNA  
 Residues: 1-8 <RES>

Cross-references: GB:M64481; NID:9181617; P1DN:AAA52318.1; P1D:9181618  
 Hsi, K.L.; Copeland, W.C.; Wang, T.S.F.

Nucleic Acids Res. 18, 6231-6237, 1990  
 Title: Human DNA polymerase alpha catalytic polypeptide binds ConA and RCA and contain  
 Reference number: S12665; MUID:91057099; PMID:2243771

Accession: S12665  
 Molecule type: protein

Residues: 19-37; 'C', 1406-1425, 'C' <HSI>  
 Genetics:

Gene: GDB:POLA  
 Cross-references: GDB:120304; OMIM:312040

Map position: Xp22.3-Xp21.1  
 Superfamily: DNA polymerase

Keywords: DNA binding; DNA replication; nucleotidyltransferase; zinc finger  
 654-691/Region: zinc finger CHCC motif  
 1249-1374/Region: zinc fingers

Query Match 5.0%; Score 88.5; DB 1; Length 1462;  
 Best Local Similarity 20.3%; Pred. No. 42;

Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

Y 16 KRDFRPKPDYCYQAKYTPCPTG-----SPIVWEGD-----DIEV-----FRL 55  
 b 183 KRKRSIGASPNPFSVTATAVPSGKIASPVSRKEPLTVPPLKRAFPADDDQVESTEEQ 242  
 Y 56 QAPVEEYKGDLLGLKIMHDAIGRSLTIGKNYTMVEYELFQLCNCFPHLRPMADP 116  
 b 243 ESGAMEFEDGDP-----DEMEVEEV-----DLEP-MAKA 272  
 Y 116 WCNQACAFCEFGIDVHKNEN---GLIVQVATISGNMFMQAKM-VKQDNETGIYETWN 171  
 b 273 WDKBS-----EPAEYVKQEADSGKGTV-----SYLGSFLPDVSCMDIDEGSSRSVQEVQ 323  
 Y 172 VKASP-----EKQAEI-----WFDSDYD-----CSKVFYLTFFNKLAIEFGAE-----FNTI 210  
 b 324 VDSSTPLVKGADBEQVHFHFWLDAVEQYQPGVVFLEFGKWLBSAETHVSCVMVKN 383

QY 211 ETTYRTIFLYSGEPYTLGNETSVPFGTGNKT-LGLAKRFYPPFKPHLPTEKELLSLQI 269  
 Db 384 ERT-----LY-----FLREMKIDINTGKETGTPISMDVHEEFPEKATKYIMK---- 429  
 QY 270 FDAVIVHKQFYLFFNFEYWFLLPMKFPFIKITYE-----EIPLPINKTUS 314  
 Db 430 FRSKVEKN-----YAFETPDVPEKSEYILEVKSAMRQLPOLDKETFS 474

## RESULT 13

D69459

Conserved hypothetical protein AF1677 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 17-Mar-2000

Accession: D69459  
 R.Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
 J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997

A/Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Arriach, P.; Kaine, B.P.; Sykes, S.A.  
 Smith, H.O.; Weese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: D69459  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA  
 A/Residues: 1-411 <KLE>

A/Cross-references: GB:A6000987; GB:A600782; NID:92689310; P1DN:AAB89568.1; P1D:9264887;  
 C/Superfamily: hypothetical protein H10701

Query Match 5.0%; Score 88; DB 2; Length 411;  
 Best Local Similarity 20.3%; Pred. No. 9.2;

Matches 55; Conservative 45; Mismatches 97; Indels 74; Gaps 13;

QY 44 MEGDDIEVFRLQAPVWEKYGDLGLHKIMHDAIG-----FSTLTGKNYTMVEYL 96  
 Db 38 LSEDDDFLIIRVEKNW-----DTLNFARVLSNMLGISQKISFAGTCDKRALTYQYFEI 92  
 QY 97 FQLNGCTFPHLRPEKADPFCNQGACFFEGIDVHMK-----NGTLVQVATISGN 148  
 Db 93 YGVKK-----EEIERVNLDAKIEVIGARRAQLQGLDLGN 128  
 QY 149 MENQAKVVKQDNENGIYETWNVAKSPKGAEMF--DSVDCSEVARTFNKLAIEFGAE 206  
 Db 129 FF-RIRYVCGRDB--LFOETRN--ELMEKGTPEFGLQRFQSTIRFTHEVGKLI----- 178  
 QY 207 FKNIEIYTRIF-LYSGEPYTLGNE-----TSVFGPTGNKTLGLAKRFYPPFKPHLP 260  
 Db 179 ---LQNNVEAFWYVAKFPFEGNEBVRKIREILMEPRDALTGLRELPKYLRVENLLOK 235  
 QY 261 -----EFLLSLLQIFDAVIVHK-QYLYF 282  
 Db 236 LREKSEEEALLSLPKNLKMMFVAYQSYIF 266

## RESULT 14

S27263

synaptic vesicle protein SV2 - bovine

N/Alternate names: transporter-like protein p87  
 C/Species: Bos primigenius taurus (cattle)

C/Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 18-Aug-2000  
 C/Accession: S27263

R.Gingrich, J.A.; Andersen, P.H.; Tiberti, M.; El Mestikawy, S.; Jorgensen, P.N.; Fremean  
 FEBS Lett. 312, 115-122, 1992

A/Title: Identification, characterization, and molecular cloning of a novel transporter-1  
 A/Reference number: S27263; MUID:93050176; PMID:1426240

A/Accession: S27263  
 A/Molecule type: mRNA

A/Residues: 1-742 <GIN>  
 A/Cross-references: GB:S47919; NID:9259173; P1DN:AAB24028.1; P1D:9259174  
 C/Superfamily: synaptic vesicle protein SV2

Query Match 5.0%; Score 88; DB 2; Length 742;  
Best Local Similarity 22.1%; Pred. No. 19;  
Matches 57; Conservative 36; Mismatches 97; Indels 68; Gaps 14;

```
ly      86 GKNYTWEMVTLFQIGCTFPH-----LRPMADP-FMCMQSG---AACFEGIDVPHMK 134
      325 GSAYQPSHSMVFWL-VCAFPVFAIGALTTPQPSRPFLENGHDEAMVYLKQVHDTNMR 383
ly      135 ENGLTVQVATISG-NMFMQAKMVKODNETGIYYETMNVKASPEKGAETMPDSDCKFV 193
      384 AKGHPREVSVTIKTIHQDELEIQTSDTGAHYQKGVKRAL-SLGGQVKNFLSC----- 438
ly      194 LRTFNKLAEFGAEFKNIETN-----YTRIFLYSGEPTVLGN-----ETSVF-- 234
      439 -----FGEYRRTITLMMGWFTMSRSGYGLTWMPFDMIRHLQAVDYARTKVPFG 489
ly      235 ---GPTGKTLGLAKRFPYPPKP-----HLPTKEFLLSLQ--IFDAVIYHKQFY--- 283
      490 ERVEHTVFNTLENOIHRGGQYFNDRKFGIRLKSVSFEDSLFECEYFEDVTSNTFFRNC 543
ly      281 ----LFTN---PEYWFL 290
      550 TFINVVFYNTDLPEYKGV 567
ly      550
```

## RESULT 15

```
71917
:okin-like outer membrane protein jhp0556 - Helicobacter pylori (strain J99)
:Species: Helicobacter pylori
:Variety: strain J99
:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
:Accession: D71917
:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
:ature, 397, 176-180, 1999
:title: Genomic sequence comparison of two unrelated isolates of the human gastric path
:Reference number: A71800; KUID:99120557; PMID:9923682
:Accession: D71917
:Status: preliminary
:Molecule type: DNA
:Residues: 1-3194 <ARN>
:Cross-references: GB:AE001488; GB:AE001439; NID:94155100; PIDN:AA06134.1; PID:9415510
:Experimental source: strain J99
:Genetics:
:Gene: jhp0556
```

Query Match 5.0%; Score 88; DB 2; Length 3194;  
Best Local Similarity 23.2%; Pred. No. 1.2e+02;  
Matches 58; Conservative 29; Mismatches 69; Indels 94; Gaps 15;

```
ly      80 FRETLLGKNVYTMWYMLFQIGCTFPHLRPMADPFCN-----QGACFEGIDVPHMK 134
      513 FAQTYGKKSALVF-----NATTP-----MANGALPKSNSYTRFGYEGVNWG 555
ly      135 ENGLTVQVAT-----ISGNMFQMAKMKVODNETGIYYETMNVKASPEKGAETMPDSDYC 189
      556 KTGVIINGTFTADRVYITGNMS-----GNGAQGTGGA--T 588
ly      190 SKFVLTFENKLAFFGAPFKNIETNYTRIFLYSGEPTY--LGNETSVPGPTGNKTLGLAIK 247
      589 LNFVGAETEINIA--GATFKNLKTTSONSYW-----TFMALGN-----GSGSGKINVSOS 635
ly      248 RFY-----YFEPKHLPTKRELLSLQIFDAVIYHKQFYLF-----YNFEYV-FIPMKF 294
      636 DFDYDMTDGSDGIDFTGN-----GVFDSVNFKAHYKFGAENSYNFKNTNPLAGNF 684
ly      295 PFI-KITYEE 303
      685 KFOGKTTIEK 694
```

Search completed: June 8, 2004, 14:08:10  
Job time: 23.8222 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

\* protein - protein search, using sw model

on: June 8, 2004, 13:56:38 ; Search time 13.1667 seconds

(without alignments)  
1249.684 Million cell updates/sec

File: US-10-010-050a-2\_COPY\_31\_346  
Perfect score: 1758  
Sequence: 1 RVSGIPSRHWPVYKRRDF.....IKITYEIPLPTRNKLTSL 316

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Search: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1758	100.0	407	1	CLN5_HUMAN
2	92	5.2	502	1	OXNA_AQUAE
3	90	5.1	585	1	NEP2_TREUV
4	89.5	5.1	427	1	ARP3_ACAQA
5	88.5	5.0	1462	1	DPOA_HUMAN
6	88	5.0	411	1	TRUD_ARCTU
7	87.5	5.0	755	1	P100_HSVJ
8	87	4.9	1184	1	CHSD_EMENT
9	85	4.8	254	1	SSA2_MOUSE
10	83.5	4.7	2167	1	BPH1_YEAST
11	83	4.7	228	1	YC01_ARCTU
12	83	4.7	889	1	CI22_HUMAN
13	83	4.7	999	1	HGPB_HAETN
14	83	4.7	1028	1	FDXG_HAETN
15	83	4.7	1453	1	APR_HUMAN
16	82.5	4.7	564	1	TM16_HUMAN
17	82.5	4.7	672	1	ACSA_METSO
18	82	4.7	2492	1	ATRX_HUMAN
19	82	4.7	2492	1	ATRX_PANTR
20	81.5	4.6	514	1	CHB2_SCHPO
21	81.5	4.6	606	1	VE1_HPV7
22	81	4.6	644	1	URCM_RAT
23	81	4.6	1451	1	DPOA_RAT
24	80.5	4.6	282	1	HCHA_ECO57
25	80.5	4.6	544	1	OXNA_BORBU
26	80.5	4.6	555	1	POLG_DENIT
27	80.5	4.6	738	1	SEC6_DROME
28	80.5	4.6	835	1	VW3_ROTSL
29	80	4.6	509	1	APPI_ORYSA
30	80	4.6	2492	1	ATRX_PONTR
31	79.5	4.5	511	1	STPA_METH
32	79.5	4.5	603	1	VE1_HPV8
33	79	4.5	508	1	ASPR_HORVU

34	79	4.5	682	1	TDR5_HUMAN	Q8nat2 homo sapien
35	79	4.5	928	1	ODO1_RICCN	Q92j42 rickettsia
36	79	4.5	2292	1	POLG_EMCVB	P17593 e genome po
37	79	4.5	2292	1	POLG_EMCVB	P17594 e genome po
38	78.5	4.5	218	1	VIXS_BPK3	P10393 bacteriopia
39	78.5	4.5	282	1	HCHA_ECOL6	P59331 escherichia
40	78.5	4.5	282	1	HCHA_ECOL1	P31658 escherichia
41	78.5	4.5	282	1	HCHA_SHIFL	P59332 shigella fl
42	78.5	4.5	643	1	RHNI_MOUSE	Q61085 mus musculu
43	78.5	4.5	660	1	AMY_BACST	P00691 bacillus su
44	78.5	4.5	1764	1	YK84_YEAST	P34241 saccharomyc
45	78.5	4.5	2358	1	MOXD_SCHPO	Q9y719 schizosach

## ALIGNMENTS

RESULT 1	CLN5_HUMAN	STANDARD	PRT	407 AA.
AC	075503:			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Ceroid-lipofuscinosis neuronal protein 5 (CLN5 protein).			
GN	CLN5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A., VARIANT VLINCL ASN-279, AND VARIANT ARG-368.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=98324783; PubMed=9662406;			
RA	Savukoski M., Klockars T., Holmberg V., Santavuori P., Lander E.S.,			
RA	Peltonen L.;			
RT	"CLN5, a novel gene encoding a putative transmembrane protein mutated			
RT	in Finnish variant late infantile neuronal ceroid lipofuscinosis.";			
RL	Nat. Genet. 19:286-288(1998).			
RN	(2)			
RP	SUBCELLULAR LOCATION, AND GLYCOSYLATION.			
RX	MEDLINE=21968572; PubMed=11971870;			
RA	Isoomppi J., Vesä J., Jalanko A., Peltonen L.;			
RT	"Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5			
RT	protein.";			
RL	Hum. Mol. Genet. 11:885-891(2002).			
CC	- FUNCTION: Not known.			
CC	- SUBCELLULAR LOCATION: Lysosomal.			
CC	- TISSUE SPECIFICITY: Ubiquitous.			
CC	- PTM: Glycosylated.			
CC	- DISEASE: Defects in CLN5 are the cause of Finnish variant late-			
CC	infantile neuronal ceroid lipofuscinosis (VLINCL) [MTM:256731];			
CC	also known as ceroid lipofuscinosis neuronal 5 (CLN5). VLINCL is a			
CC	fatal childhood neurodegenerative disease characterized by			
CC	progressive visual and mental decline, motor disturbance, epilepsy			
CC	and behavioral changes. The first symptom is motor clumsiness,			
CC	followed by progressive visual failure, mental and motor			
CC	deterioration and later by myoclonia and seizures.			
CC	- DATABASE: NAME=NCL CLN5;			
CC	NOTE=Neural Ceroid Lipofuscinoses mutation db;			
CC	WWW="http://www.ucl.ac.uk/ncl/CLN5.html".			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF068227; AAC27614.1; --			
DR	GeneW; HGNC:2076; CLN5.			
DR	MIM; 608102; --			

```

MIM; 256731; -.
GO; GO:0016021; C: integral to membrane; TAS.
GO; GO:0008151; P: cell growth and/or maintenance; TAS.
M Transmembrane; Lysosome; Glycoprotein; Neuronal ceroid lipofuscinosis;
M Disease mutation; Polymorphism; Epilepsy.
M TRANSMEM 75 91
M CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
M CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
M VARIANT 279 279 D -> N (in VLINC).
M /FTID=VAR_005137.
M VARIANT 368 368 K -> R (in dSNP:1800209).
M /FTID=VAR_005138.
M SEQUENCE 407 AA; 46339 MW; 449702D1DC9BFEB4 CRC64;

Query Match 100.0%; Score 1758; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 RVSGIPSRHMPVYKRFDPKDPYCAKYTCFSGPIPMWEGDDIEVRLQAPY 60
Y 92 RVSGIPSRHMPVYKRFDPKDPYCAKYTCFSGPIPMWEGDDIEVRLQAPY 151
Y 61 EFKYGLGLGLKIMHDAIGRSTLTGKNYMEYELFOLGNCFFPHLRPMMDAPFMCNG 120
Y 152 EFKYGLGLGLKIMHDAIGRSTLTGKNYMEYELFOLGNCFFPHLRPMMDAPFMCNG 211
Y 121 AACFFEGIDVHWKENGTLVQVATISGNMKNQAKVKNODNETGIYETNNVASEPKA 180
Y 212 AACFFEGIDVHWKENGTLVQVATISGNMKNQAKVKNODNETGIYETNNVASEPKA 271
Y 181 EFWEDSDCKEFLVRFNKLAEFGAEFKNIETNYTRIFLYSGEPTLYGNETSYGCEPTGNK 240
Y 272 EFWEDSDCKEFLVRFNKLAEFGAEFKNIETNYTRIFLYSGEPTLYGNETSYGCEPTGNK 331
Y 241 TLGLAKRFYPRKPHLPTKEFLSLIOTFDVAVIHKQFLPNEFYWFLPMKEPPIKTT 300
Y 332 TLGLAKRFYPRKPHLPTKEFLSLIOTFDVAVIHKQFLPNEFYWFLPMKEPPIKTT 391
Y 301 YEELPIRKTKLSGL 316
Y 392 YEELPIRKTKLSGL 407

RESULT 2
XXAA_AQVAF STANDARD; PRT; 502 AA.
ID XXAA_AQVAF STANDARD; PRT; 502 AA.
OC 06561;
Y 15-MAR-2004 (Rel. 43, Created)
Y 15-MAR-2004 (Rel. 43, Last sequence update)
Y 15-MAR-2004 (Rel. 43, Last annotation update)
M Inner membrane protein oxaa.
M OXAA OR AQ_175.
M Aquifex aeolicus.
M Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
M NCBI_TaxID=63363;
M [1]
M SEQUENCE FROM N.A.
M STRAIN=VF5;
M MEDLINE=98196666; PubMed=9537320;
M Deckert G., Warren P.V., Gaasterland T., Young W.G., Tenox A.L.,
M Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
M Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
M "The complete genome of the hyperthermophilic bacterium Aquifex
M aeolicus";
M Nature 392:353-358(1998).
M -!- FUNCTION: Required for the insertion of integral membrane proteins
M into the membrane. Probably plays an essential role in the

```

```

CC integration of proteins of the respiratory chain complexes.
CC Involved in integration of membrane proteins that insert
CC independently and independently of the Sec translocase complex (By
CC similarity).
CC -!- SUBUNIT: Specifically interacts with transmembrane segments of
CC nascent integral membrane proteins during membrane integration (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the OXA1/OXA family. Subfamily 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000677; AAC06534.1; -.
CC PIR; F70316; P70316.
CC HAMAP; MF_01810; -.
CC DR InterPro; IPR001708; 60kDa_innermem.
CC DR Pfam; PF02096; 60KD_IMP; 1.
CC KM Transmembrane; Inner membrane; Complete proteome.
CC FT TRANSMEM 12 34 POTENTIAL.
CC FT TRANSMEM 308 330 POTENTIAL.
CC FT TRANSMEM 378 400 POTENTIAL.
CC FT TRANSMEM 447 469 POTENTIAL.
CC SQ SEQUENCE 502 AA; 58224 MW; 6C24B8405F2570AD CRC64;

Query Match 5.2%; Score 92; DB 1; Length 502;
Best Local Similarity 21.3%; Pred. No. 2.2; Indels 68; Gaps 15;
Matches 63; Conservative 51; Mismatches 114; Indels 68; Gaps 15;

Q 40 P.PWEGDDIEVRLQAPVWEFKYGLGLKIMHDAIGRSTLTGKNYMEYELFOL 99
Q 118 PLFIPTGNLDLQ-KLNFGEYEIKKK--NSVELHKKELKVKKILSYKGA!----- 166
Q 100 GNCFFPHLRPE-MDAPFMCNGA---ACFFEGIDVHWKENGTLVQVATISGNMKNQAK 154
Q 167 -----HLSVEGLKPFYFWFVGSPPDEAFYTHGVV-LKINGEVVRLDV----- 209
Q 155 KMWKODNETGIYETNNVASEPKAETWEDSDCKEFLVRFNKLAEFGAEFKNIETNY 214
Q 210 -----DGLKINFEBSNIFGGBESRYFFKGAQDYKAIYK---VKLGDFVSLST-- 258
Q 215 TRIFLYSGEPT-YLG-----NETSVFGPTGNKTLGLAIK--RFYFPKPHLPTKEPL 263
Q 259 ---FLYDGEKTYLGLAKQYARLRRLGLVDTLDMGTLKIIVKPLFLVLYIEH--TGSWV 313
Q 264 LSL-----QTFDAVIVHKQFLPNEFYWFLPMKEPPIKITYEELPIRKTKL 313
Q 314 LSLVLTFTVRIEFLPLGYKSVSMOKLOE--LAPMKRKKKXKDDPVKQDEMM 367

RESULT 3
NEP2_THEVU STANDARD; PRT; 585 AA.
ID NEP2_THEVU STANDARD; PRT; 585 AA.
AC Q08751;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neopullulanase 2 (EC 3.2.1.135) (alpha-amylase II) (TVA II).
GN TVAIL.
OS Thermoactinomyces vulgaris.
OC Bacteria; Firmicutes; Bacillales; Thermoactinomycetaceae;
OC Thermoactinomycetes.
OC NCBI_TaxID=2026;
OC [1]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP STRAIN=R-47;
RC MEDLINE=93222535; PubMed=7763540;
RX

```

Tonozuka T., Ohtsuka M., Mogi S.-I., Sakai H., Ohta T., Sakano Y.,  
"A neopululanase-type alpha-amylose gene from Thermactinomyces  
vulgaris R-47.";  
Biosci. Biotechnol. Biochem. 57:395-401(1993).  
[2]  
X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
STRAIN=R-47;  
MEDLINE=99241045; PubMed=10222200;  
Kamitori S., Kondo S., Okuyama K., Yokota T., Shimura Y., Tonozuka T.,  
Sakano Y.,  
"Crystal structure of Thermactinomyces vulgaris R-47 alpha-amylose II  
(TVAII) hydrolyzing cyclodextrins and pullulan at 2.6-A resolution.";  
J. Mol. Biol. 287:907-921(1999).  
[3]  
X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
MEDLINE=22047855; PubMed=12051850;  
Kamitori S., Abe A., Ohtaki A., Kaji A., Tonozuka T., Sakano Y.,  
"Crystal structures and structural comparison of Thermactinomyces  
vulgaris R-47 alpha-amylose I (TVAI) at 1.6 A resolution and  
alpha-amylose 2 (TVAII) at 2.3 A resolution.";  
J. Mol. Biol. 318:443-453(2002).  
-!- FUNCTION: Hydrolyzes pullulan efficiently but only a small amount  
of starch. Endohydrolysis of 1,4-alpha-glucosidic linkages in  
pullulan to form panose. Cleaves also (1-6)-alpha-glucosidic  
linkages to form maltotriose.  
-!- CATALYTIC ACTIVITY: Hydrolysis of pullulan to panose (6-alpha-D-  
glucosylmallose).  
-!- COFACTOR: Birds 1 calcium ion per subunit.  
-!- SUBUNIT: Monomer.  
-!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D13178; BA002473.1; -.  
PIR; JCI486; JCI486.  
PDB; 1BVZ; 02-MAR-99.  
PDB; 1G1Y; 14-MAR-01.  
PDB; 1JF5; 25-DEC-02.  
PDB; 1JF6; 25-DEC-02.  
PDB; 1J12; 18-DEC-02.  
PDB; 1J1B; 12-AUG-03.  
PDB; 1J18; 18-DEC-02.  
InterPro; IPR006589; Alp\_amy1\_cat\_sub.  
InterPro; IPR006047; Alpha\_amy1\_cat.  
InterPro; IPR004185; Glyco\_hydro\_13ig.  
InterPro; IPR007110; 19-like.  
Pfam; PF00128; alpha-amylose; 1.  
Pfam; PF02903; alpha-amylose\_N; 1.  
SMART; SMO0642; Amy; 1.  
Carbohydrate metabolism; Hydrolase; Glycosidase; Calcium-binding;  
3D-structure  
ACT\_SITE 325 325  
ACT\_SITE 354 354  
ACT\_SITE 421 421  
METAL 143 143 CALCIUM.  
METAL 143 143 CALCIUM (VIA CARBONYL OXYGEN).  
METAL 148 148 CALCIUM.  
METAL 149 149 CALCIUM.  
METAL 169 169 CALCIUM (VIA CARBONYL OXYGEN).  
METAL 171 171 CALCIUM.  
HELIX 3 5  
STRAND 7 7  
TURN 12 14  
STRAND 15 19  
TURN 20 21  
STRAND 22 30  
TURN 31 32

FT STRAND 36 42  
FT TURN 44 45  
FT TURN 48 49  
FT STRAND 53 62  
FT STRAND 66 74  
FT TURN 76 77  
FT STRAND 80 87  
FT STRAND 93 97  
FT TURN 98 99  
FT STRAND 100 102  
FT HELIX 105 108  
FT STRAND 111 113  
FT TURN 119 120  
FT HELIX 126 130  
FT STRAND 133 136  
FT HELIX 138 141  
FT TURN 146 147  
FT TURN 151 152  
FT STRAND 153 153  
FT TURN 158 159  
FT TURN 164 165  
FT STRAND 167 167  
FT HELIX 172 185  
FT TURN 186 186  
FT STRAND 189 192  
FT STRAND 196 196  
FT STRAND 212 212  
FT TURN 214 216  
FT HELIX 219 230  
FT TURN 231 233  
FT STRAND 236 240  
FT TURN 242 243  
FT STRAND 244 244  
FT TURN 247 248  
FT HELIX 250 251  
FT HELIX 259 261  
FT TURN 263 264  
FT HELIX 265 267  
FT STRAND 268 268  
FT STRAND 270 270  
FT STRAND 283 283  
FT STRAND 285 285  
FT TURN 291 292  
FT STRAND 294 296  
FT TURN 298 299  
FT HELIX 301 317  
FT TURN 318 318  
FT STRAND 321 324  
FT TURN 325 326  
FT HELIX 327 329  
FT TURN 332 345  
FT STRAND 347 348  
FT STRAND 350 353  
FT TURN 360 363  
FT STRAND 370 376  
FT STRAND 374 384  
FT HELIX 385 386  
FT TURN 391 402  
FT TURN 403 404  
FT HELIX 407 410  
FT TURN 411 412  
FT STRAND 414 416  
FT TURN 420 421  
FT HELIX 425 428  
FT TURN 429 431  
FT HELIX 433 443  
FT TURN 444 445  
FT STRAND 449 453  
FT TURN 454 455  
FT HELIX 456 458  
FT TURN 459 459  
FT TURN 465 468  
FT HELIX 476 478

	HELIX	481	496	
T	TURIN	498	498	
T	HELIX	499	502	
T	STRAND	504	511	
T	TURIN	512	515	
T	STRAND	516	523	
T	TURIN	524	525	
T	STRAND	526	533	
T	STRAND	539	544	
T	STRAND	553	555	
T	STRAND	561	561	
T	STRAND	564	565	
T	TURIN	566	567	
T	STRAND	568	573	
T	TURIN	575	576	
T	STRAND	578	582	
HQ	SEQUENCE	585 AA;	67467 MW; E311813A05A7791A CRC64;	
	Query Match	5.1%;	Score 90;	DB 1;
	Best Local Similarity	21.58;	Pred. No. 4.1;	Length 585;
	Matches	53;	Conservative	30; Mismatches 68; Indels 96; Gaps 15;
IY	IKIILDAVFHAGDQFFAFPDVLQKGQSXYKMFIEDPEVSTKTSTNVEFAYQVPAM	234		293
IY	71 LKIMHDAI-----GPRSTLT--GKNYTMEW-----YELFOLGNCTF	104		143
IY	105 PHLR---PEMD-----APFWCNOGAACFEFG-----IDVHKMENGTLVGCV---	A		143
IY	294 PKLRTPEVEXKEYLFVDVARFEMEG-----IDGWELDVANVEDHAFMREFRLLVKSLNPDA	349		409
IY	144 TIGGNMFNQAKKKVKKQNGTGI-----YYEFWNVKAS-----PEKG	179		230
IY	350 LIAGEIHWHDSSGMGLMDQFDPSVMNYLFRESVIRPFATGEIHAEFPDAELTEARMLYEBOA	409		457
IY	180 AE-TW---FDSYDOSKFPLTRTFNKLAERGAERKNIETMYTRI-----FLYSGEP-TYIGNE	230		457
IY	410 AQLGMLMLDSHTDFLTRFLTSCGGNEAKF-----RLAVLPOMTYLSTPIITYGDE	457		457
IY	231 TSVFGPT	237		457
IY	458 IGWAGAT	464		457
RESULT 4				
ID	ARP3 ACACA	STANDARD;	PRT;	427 AA.
NC	PS3480:			
YT	01-OCT-1996 (Rel. 34, Created)			
YT	01-OCT-1996 (Rel. 34, Last sequence update)			
YT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Actin-like protein 3.			
IN	ARP3.			
NS	Acanthamoeba castellanii (Amoeba).			
YC	Eukaryota; Acanthamoebidae; Acanthamoeba.			
UN	NCBI_TaxID=5755;			
UP	SEQUENCE FROM N.A.			
UC	STRAIN=ATCC 30010 / Neff;			
UC	MEDLINE=66017709; PubMed=7593166;			
UA	Kellner J.F., Atkinson S.J., Pollard T.D.;			
UT	"Sequences, structural models, and cellular localization of the			
UT	actin-related proteins Arp2 and Arp3 from Acanthamoeba.";			
JL	J. Cell Biol. 131:385-397(1995).			
XC	-!- FUNCTION: Part of a complex implicated in the control of actin			
XC	polymerization in cells (By similarity).			
XC	-!- SUBUNIT: Belongs to a complex composed of ARP2, ARP3, P41-ARC,			
XC	P34-ARC, P21-ARC, P20-ARC and Pl6-ARC (By similarity).			
XC	-!- SIMILARITY: Belongs to the actin family. ARP3 subfamily.			
XC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
XC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
XC	the European Bioinformatics institute. There are no restrictions on its			
XC	use by non-profit institutions as long as its content is in no way			
C	contradict the original source.			

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U29610; AAA93068.1; ".
DR InterPro; IPRO0400; Actin_like.
DR Pfam; PF00022; actin; 1.
DR SMART; SMD0268; ACTIN; 1.
DR PROSITE; PS0113; ACTINS_ACT_LIKE; 1.
DR Structural Protein; Cytoskeleton.
SQ SEQUENCE 427 AA; 48635 MW; A53A0965B1E7EAP6 CRC64;

Query Match 5.1%; Score 89.5; DB 1; Length 427;
Best Local Similarity 20.6%; Pred. No. 3;
Matches 64; Conservative 49; Mismatches 96; Indels 101; Gaps 17;

QY 25 DPYCAKXTCPC--TGSPPIVME---GDDIE--VFRLQAPWEFKYGDLLGHLKIMHD 76
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 44 DPCRARRRMWCPMAAGNIADLDFPIGDEAYENSKYQITMPV--RHGV----- 91
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 77 AIGRSTLTGKNYT-MEWYELFOLGNTCPHLR-----PEMDAPMCOGAAC 123
   :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 92 -----ENWTHMEQF---WEHCIFXKRCPEDDHFFLLTEPLNAPENREYAEI 137
   :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 124 PFEGIDVHAKENGLVQVATISGNFNQAKVYKQDNFGITYETW----- 170
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 138 MFE-----TNNVGLYIAQAVLALAAWTSKQVTEKTLTGTVIDSGDVTHTVIPAEGY 192
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 171 ----NVAKSPKAE--TWF-----DSYDCKFVLRTPENKLA-ERGAEFKN 209
   ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 193 VIGSSIHITPLAGRDITNFTQLLRNENKIPPAETIEVAKRIKETISYCPDLYVEFK 252
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 210 IETNYRIFLYSGEPTLYLGNETSVPFGPTGNK--TLGLAIRFYYP--FKPHLPTEKELL 264
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 253 YDEPDKWF-----KTYEGIES-----VGKKPVNVQVGYERFLGPELFFNPEIFSSPFLT 302
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 265 SLQIIPDAVI 274
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 303 PLPKVDETL 312
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 5
DPOA HUMAN STANDARD; PRT; 1462 AA.
AC P09884;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7) .
GN POLA.
OS Homo sapiens (Human) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88136090; Pubmed=3359994;
RA Wong S.W., Wahl A.F., Yuan P.-M., Arai N., Pearson B.E., Arai K.,
RA Korn D., Hunkapiller M.W., Wang T.S.-F.;
RT "Human DNA polymerase alpha gene expression is cell proliferation
RT dependent and its primary structure is similar to both prokaryotic
RT and eukaryotic replicative DNA polymerases.";
RL EMBO J. 7:37-47(1988) .
RN [2]
RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE=91172197; Pubmed=2005899;
RA Pearson B.E., Nashner H.P., Wang T.S.;
RT "Human DNA polymerase alpha gene: sequences controlling expression in
RT cycling and serum-stimulated cells.";
RL Mol. Cell. Biol. 11:2081-2095(1991) .
CC -1- FUNCTION: Polymerase alpha in a complex with DNA primase is a
CC replicative polymerase.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate

```

```

+ (DNA) (N).
-1- SUBCELLULAR LOCATION: Nuclear.
-1- MISCELLANEOUS: In eukaryotes there are five DNA polymerases:
    alpha, beta, gamma, delta, and epsilon which are responsible for
    different reactions of DNA synthesis.
-1- SIMILARITY: Belongs to the DNA polymerase type-B family.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL: X06745; CAA29920.1; -
EMBL: M64461; AAA52318.1; -
PIR: S00257; DJHDUC.
Genew: HGNC:9173; POLA.
R GK: P09884; -
R MIM: 312040; -
R GO: GO:0005634; C:nucleus; NAS.
R GO: GO:0003889; F:alpha DNA polymerase activity; NAS.
R GO: GO:0006260; P:DNA replication; NAS.
R InterPro: IPR006172; DNA_pol_B.
R InterPro: IPR006134; DNA_pol_B_dom.
R InterPro: IPR006133; DNA_pol_B_exo.
R InterPro: IPR004578; Pol2.
R Pfam: PF00136; DNA_pol_B_1.
R Pfam: PF03104; DNA_pol_B_exo; 1.
R PRINTS: PR00106; DNAPOLB.
R SMART: SM00486; POLBc; 1.
R TIGRFAMs: TIGR00592; pol2; 1.
R PROSITE: PS00116; DNA_POLYMERASE_B; 1.
R Transferrase; DNA-directed DNA polymerase; DNA replication;
W DNA-binding; Nuclear protein.
W DNA BIND 650 715 POTENTIAL.
T DNA BIND 1245 1376
Q SEQUENCE 1462 AA; 165860 MW; 25C270B0A0DB38BE CRC64;

Query Match 5.0%; Score 88.5; DB 1; Length 1462;
Best Local Similarity 20.3%; Pred. No. 17;
Matches 71; Conservative 56; Mismatches 115; Indels 107; Gaps 20;

16 KRFDPRPKDPCQAKYKPCPTG-----SPPVWEGD---DRIEV-----FRL 55
183 KRRTGASNPSPSVHTATAVPSSGKTASPVSRKPEPLTPPLRAEPAGDVGVSTEEBQ 242
56 QAPVEFKYGDLLGHUKIMHDAIGFRSLTGKNTYMEWYELFQLGNCETPEHLRPMEDAPF 115
243 ESGAMEFEGD-----DERMEVEV-----DLEP-MAKA 272
116 WCNQAGACEFEGIDVHWKEN--GTLVQVATISGMENQAKW-VKQNETGIYETWN 17.
273 WKES-----EPAEVKKQADSGKGTV-----SYLGSFLPDVSCWPDIDGDSFSVQEVQ 323
172 VQASP-----EKAGET-----WDSYD---CSKPVLRPNKLAEGAE-----FNKI 210
324 VDSHLPLVKAGADEQVHFHYWLDAYEDQYNQPGVVFLEKWMISAEHVSQCVWVKNI 383
211 ENNYRIEFLYSGEPTVIGNETSVFPGTGNKT-LGLAKRFYYPFKPHLPTEKFLSLDQI 269
384 ERT-----LY-----FLPREMKIDINTGKGTGTPISMKQDVVEEPDEKATKKIKK- 429
270 FDVAIVHQOYLFYNPEYWFLEPMKRPFIKTYE---EIPDPIRKTLS 314
430 FKSKEVEK-----YAFELPDVPEKSEYLEVKYSALMPQIPOLDKETFS 474

ESUT 6
RUD_ARCFU STANDARD; PRT; 411 AA.
D TRUD_ARCFU
C 028596;

```

```

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable tRNA pseudouridine synthase D (EC 4.2.1.70) (Pseudouridylylate
DE synthase) (Uracil hydrolase).
GN TRUD OR AF1677.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_Taxid=2234;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klelek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.U., Kertavagge A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goeyne J.D., Weidman U.F., McDonald L., Uitterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -1- FUNCTION: Could be responsible for synthesis of pseudouridine from
CC uracil-13 in transfer RNAs (By similarity).
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: Belongs to the pseudouridine synthase trnd family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE000987; AAB89568.1; -
CC PIR: D69459; D69459.
CC TIGR: AF1677; -
CC HAMAP: MF_01082; -, 1.
CC InterPro: IPR001656; UPP0024.
CC Pfam: PF01142; UPP0024; 1.
CC TIGRFAMs: TIGR00094; TIGR00094; 1.
CC PROSITE: PS01268; UPP0024; 1.
CC tRNA processing; Lyase; Complete proteome.
CC ACT SITE 81
CC SITE 81
SQ SEQUENCE 411 AA; 47715 MW; 109B5250D969922E CRC64;

Query Match 5.0%; Score 88; DB 1; Length 411;
Best Local Similarity 20.3%; Pred. No. 3.8;
Matches 55; Conservative 45; Mismatches 97; Indels 74; Gaps 13;

44 MEGDDIEVFRLQAPVMEFKYGDLLGHUKIMHDAIG-----FSLTGKNTYMEWYEL 96
38 LSDSDGFLIIRVEKKNM-----DTLNFARVLISNALGISQKISFAGTQDKALTVQYPSI 92
97 FQLGNCETPEHLRPMEDAPFCWQAGACFEFGIDVHWKE-----NGTLVQVATISGM 148
93 YGVKK-----EELERNLDAKIEVIGARRALQLQDGLGN 128
149 MENOAKWVKQNETGIYETWNVVASPEKGAETWF--DSYDCSFVLRTFNKLAEGAE 206
129 FF-RIVVGCGRDGE--IFQETRN--ELMEKGPNNFGIQRFGSIRFIITHEWKKL----- 178
207 FKNITNTYTRIF-LYSGEPTVIGNE-----TSVPGTGNKTLGLAKRFYYPFKPHLPK 260
179 ---LQNNYEAFVYVAKPFEGENEVEVKRIEILMETDARKLGLRELPRKYLRERNLQK 235

```

```

Y 261 -----EFLLSLQIFDAVIVHK-QEYLF 282
| | | | | : | | | | |
b 236 LRGRKSEELALSLPRNLKMMFHAHQSYIF 266

RESULT 7
100 HSV7J STANDARD; PRT; 755 AA.
C P52519;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E large structural phosphoprotein homolog (PPI100).
U1.
S Human herpesvirus (type 7 / strain J1) (HHV7).
C viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Simplexvirus.
X NCBI_TaxID=57278;
N [1]
P SEQUENCE FROM N.A.
A Nicholas J.;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
C -1- SIMILARITY: TO THE LARGE STRUCTURAL PROTEOPROTEINS OF HSV-6 AND
HCWV U132.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
R EMBL: U43400; AAC54672.1; -.
R PIR: T41912; T41912.
W Matrix protein, Phosphorylation.
Q SEQUENCE 755 AA; 86580 MW; 4083744CCCF91DA CRC64;

Query Match
Best Local Similarity 5.0%; Score 87.5; DB 1; Length 755;
Matches 58; Conservative 37; Mismatches 98; Indels 87; Gaps 15;

Y 114 PF-WCNOGAAC-----FEEGIDID--VHMKEN-GTLVQVATISGN-----MENQAKXV 157
| | | | | : | | | | |
b 7 PRAMISDEKCKFLSRFENISSLPVDIRENPWILSQCVKNGSNINNKILYNLLIM- 65

Y 158 KQDNETGIYETWNVKASPEKGETWEDSYDSKFEV-----LRTFNKLA 202
| | | | | : | | | | |
b 66 -----IYHQTLCKKKDPDY-EEVWQELIKYOKILKDYLEQROMITDYSSILTSFNKV- 116

Y 203 FGAEFKNITNTYTRI--FLYSGEPYPLGN-----ETSVFG--PTGNKTLGLAIKRF 249
| | | | | : | | | | |
b 117 FETFEKNVAKDLKJGSLRWGVTHADYVNLTEERREIEENLOKAKONMMLSTTYOI 176

Y 250 YYPFKKH-----LPTEFLSLQD-----IFDAVIVHKQFY 280
| | | | | : | | | | |
b 177 VDPWNNNGYVNVINIRKLVLGNLLITLHGSMMMETALNTINEKNKALIKAIENNKVY 236

Y 281 LFYNEFYWFLPKK---PPFIKITYEELPLIRNKTLISGL 316
| | | | | : | | | | |
b 237 SIYSYOILSLPLTSRVTSPFKILTFEDFVITKSLIHL 276

RESULT 8
HSD EMENT STANDARD; PRT; 1184 AA.
C P78611; Q00744;
T 15-JUL-1999 (Rel. 38, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Chitin synthase D (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
E transferase D) (Class-V chitin synthase D).

```

```

GN CHSD OR CHSE.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 89;
RX MEDLINE=9628568; PubMed=8709948;
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RT "The Aspergillus nidulans genes chsA and chd encode chitin synthases
RT which have redundant functions in conidia formation."
RL Mol. Genet. 251:442-450(1996).
RN [2]
RP ERRATUM.
RX MEDLINE=97188530; PubMed=9037115;
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RL Mol. Genet. 253:520-528(1997).
RN [3]
RP REVISIONS.
RA Motoyama T., Fujiwara M., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=96406387; PubMed=8810520;
RA Specht C.A., Liu Y., Robbins P.W., Bulawa C.E., Iartchouk N.,
RA Winer K.R., Riggle P.J., Rhodes J.C., Dodge C.L., Culp D.W.,
RA Boyia P.T.;
RT "The chd and chs genes of Aspergillus nidulans and their roles in
RT chitin synthesis."
RL Funct. Genet. Biol. 20:153-167(1996).
CC -1- FUNCTION: Plays a major role in cell wall biogenesis.
CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
CC beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-
CC beta-D-glucosaminyl)}(N+1).
CC -1- SUBCELLULAR LOCATION: Plasma membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC V.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D83246; BAA1866.2; -.
DR EMBL; U52362; AAA97482.1; -.
DR InterPro; IPR004835; Chitin synth_fng.
DR InterPro; IPR001199; Cyt_B5.
DR Pfam; PF03142; Chitin synth_2; 1.
DR Pfam; PF00173; heme_1; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW Multigene family.
KW TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 476 496 POTENTIAL.
FT TRANSMEM 1039 1059 POTENTIAL.
FT TRANSMEM 1073 1093 POTENTIAL.
FT TRANSMEM 1097 1117 POTENTIAL.
SQ SEQUENCE 1184 AA; 133504 MW; D6157184F154ED5 CRC64;

Query Match
Best Local Similarity 4.9%; Score 87; DB 1; Length 1184;
Matches 75; Conservative 44; Mismatches 132; Indels 114; Gaps 20;

QY 33 TFCPGSPFVWEGDDIEVF-----RLQAPWEPKYGDLLGHLIMDAIGRS---TLT 85
| | | | | : | | | | |
DB 189 TYC---SVITTPAPDFVAKCFGMPQKQKASAREKIG-LISIIIMIAFVGSLTFGEFAT 244
| | | | | : | | | | |
QY 86 -----GKNTYMEYELFQCGNCTFFHLR--PEV-----DAPF-WCNOGAAC 123
| | | | | : | | | | |

```



```

245 VCGTPRLKINIGSGYIIFHGQAVDLTKSTHPAAAGIPMTNVLXDLPHKYGGQDSF 304
124 FEEGID-----DYHMKNGTLVQVATISGMFNQMAKVMKADNETGIYETW 170
305 PFOEVNGACKLITRTENSNDIPTNSGDLAWY--FPCAIFNDGGS--SEPIITVSYNGM 360
171 --NVKASPEKGAE-----TWFDSDSGSKFVLTFRNKLAERGAE-----PKN 205
361 ACHTSGSARKSFYSLKNSGDVYFTMEDTKNSR-----KLAVYSGVVDLNLNWFDD 413
210 IETNYTRIF-----LYSGEFTYIG--NETSVFGPTGNKTLG-LAIKR 246
414 TVNVPTFKDLRDNDIRGVDLTFYFQTGEDKQIGKCLSLIKVGSIDTDTVGCIAQV 473
249 FYYPKPKHLPTRKEFLSLQIFDAIVHKQFLFNFYFWFLPMKFPFKITVEIPIPI 306
474 VLY-----VSILFISTIVIKFAVALLFO--WFLAPFAAQKTSMGAVDSKA 516
309 RNKTL 313
519 RNQOI 523

```

RESULT 9

```

3A2_MOUSE STANDARD; PRT; 254 AA.

```

```

28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
3-oxo-5-alpha-steroid 4-dehydrogenase 2 (EC 1.3.99.5) (Steroid
5-alpha-reductase 2) (SR type 2) (5 alpha-SR2).
SRD5A2 OR 5ARR2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISUB=Kidney;
MEDLINE=21882004; PubMed=11884637;
Takeyama K., Kato S.;
"Transcriptional regulation of the mouse steroid 5alpha-reductase
type II gene by progesterone in brain.";
Nucleic Acids Res. 30:1387-1393 (2002).
-i- FUNCTION: Converts testosterone into 5-alpha-dihydrotestosterone
and progesterone or corticosterone into their corresponding 5-
alpha-3-oxosteroids. It plays a central role in sexual
differentiation and androgen physiology (By similarity).
-i- CATALYTIC ACTIVITY: A 3-oxo-5-alpha-steroid + acceptor = a 3-oxo-
delta(4)-steroid + reduced acceptor.
-i- SUBCELLULAR LOCATION: Integral membrane protein. Microsomal
intracellular membrane (By similarity).
-i- SIMILARITY: Belongs to the steroid 5-alpha reductase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB049456; BAB40179.1; -.
MGD; MGI:2150380; Str5a2.
GO; GO:0003865; F:3-oxo-5-alpha-steroid 4-dehydrogenase activity; IMP.
GO; GO:0003539; P:male genital morphogenesis; IMP.
GO; GO:0006694; P:steroid biosynthesis; IMP.
InterPro; IPR001104; Strd5A_dhc.
Pfam; PF02344; Steroid_dh; 1.
PROSITE; PS50244; SSA_REDUCTASE; 1.
Oxidoreductase; Microsome; Transmembrane; Sexual differentiation;
Multigene family.

```

```

FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
SQ SEQUENCE 254 AA; 28619 MW; 8D25019E8DC4DF47 CRC64;

```

Query Match 4.8%; Score 85; DB 1; Length 254;  
 Best Local Similarity 27.3%; Pred. No. 3.8;  
 Matches 35; Conservative 16; Mismatches 39; Indels 38; Gaps 7;

```

QY 197 FNKLAEFGAEFKNIETNY---TRIFLYSGE-PTYL-----GNETSVFGPTGNKTLGL 244
DB 27 FGKDSYSGKHSSESSGVPLPARIAWPLQELPSFVSGVWMLAQPRSLFGPPGNVLGL 86
QY 245 AIKFFYFPRKHLPTKRELSSL---QIFDAVYHK-----QYLFYNYFEY-- 287
DB 87 -----FSAHYFRTFTYSLTRGRPLSAVIFLKATAPCIINGLLQAYVLYVCAEYPE 138
QY 288 -WFLPMKF 294
DB 139 EMTDMRF 146

```

RESULT 10

```

BPH1_YEAST STANDARD; PRT; 2167 AA.

```

```

ID BPH1_YEAST
AC P25356; Q02396; Q07348;
DT 01-MAY-1992 (Rel. 22, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beige protein homolog 1.
OS BPH1 OR YCR032W OR YCR32W OR YCR591 OR YCR601.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RC MEDLINE=92116648; PubMed=1837415;
RX Rodriguez F., Martegani E., Mauri I., Alberghina L.;
RT "The sequence of 8.8 kb of yeast chromosome III cloned in lambda
PM2270 contains an unusual long ORF (YCR601).";
RL Yeast 7:631-641 (1991).
[2]
SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92244356; PubMed=1574125;
RA Oliver S.G., van der Aart O.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Bent P., Berben G., Bergantino E., Bileau N., Boille P.-A.,
RA Botolin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Cavignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doria C., Dubois E.,
RA Dujon B., Dusterhoeft A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Franchingues-Galliard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.O., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
RA Goffeau A., Grenson M., Grisevald P., Grivell L.A., Haesemann M.,
RA Hatat B., Hegemann J.H., Heibert C.J., Hilger F., Holsen S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jaquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Klehans U., Kreis P., Lafranchi G., Lewis C., van der Linden C.G.,
RA Lucchini G., Luczenkitchen K., Maat C., Mannheim P., Manzana M.E.,
RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
RA Messenguy F., Mewes H.-W., Molemans F., Montagne M.A., Navas L.,
RA Newton C.S., Olson M.V., Paillet C., Panzeri L., Pearson B.M.,
RA Pereira J., Philippsen P., Pierard A., Planza R.J., Plevani P.,
RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Reynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaffl-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala T., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,

```

A Thireos G., Triano L.N., Urrestarazu L.A., Valle G., Vetter I.,  
A van Vliet-Reedijk J.C., Volckaert G., Vreken P., Marmington J.R.,  
A von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,  
A Zimmerman F.K., Sgouros J.G.,  
A "The complete DNA sequence of yeast chromosome III.",  
L Nature 357:38-46(1992).  
L [3]  
L REVISIONS.  
L Valles G., Volckaerts G.,  
L Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
P [4]  
P SEQUENCE OF 335-2167 FROM N.A.  
P MEDLINE=91335897; PubMed=1870232;  
X Jia Y., Slonimski P.P., Herbert C.J.,  
T "The complete sequence of the unit YCS9, situated between CRY1 and  
T YAT, reveals two long open reading frames, which cover 91% of the  
T 10.1 kb segment.",  
L Yeast 7:413-424(1991).  
L [5]  
L SEQUENCE OF 1760-2167 FROM N.A.  
L MEDLINE=92133166; PubMed=1776366;  
X Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.,  
A "The complete sequence of a 7.5 kb region of chromosome III from  
T Saccharomyces cerevisiae that lies between CRY1 and MAT.",  
T Yeast 7:761-772(1991).  
L -1- FUNCTION: May be involved in acetic acid export.  
C -1- SIMILARITY: Contains 1 BRACH domain.  
C -1- SIMILARITY: Contains 4 WD repeats.  
C -----  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
C -----  
C EMBL; X62452; CAA44309.1; -  
R EMBL; X58720; CAC42982.1; -  
R EMBL; X59075; CAA41798.1; -  
R EMBL; S78624; AAB21258.1; -  
R PIR; S19444; S19444.  
R Germline; 138938; -  
R SGD; S0000628; BPH1.  
R GO; GO:0009268; P:response to pH; IMP.  
R InterPro; IPR000409; Beige\_BEACH.  
R InterPro; IPR01680; WD40.  
R Pfam; PF02138; Beach; 1.  
R ProDom; PD007848; Beige\_BEACH; 1.  
R ProSite; PS50197; BEACH; 1.  
R ProSite; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
R ProSite; PS50294; WD\_REPEATS\_REGION\_1.  
R ProSite; PS50082; WD\_REPEATS\_2; FALSE\_NEG.  
R Repeat; WD repeat.  
M REPEAT 1927 WD 1.  
M REPEAT 1976 WD 2.  
M REPEAT 2072 WD 3.  
M REPEAT 2129 WD 4.  
M REPEAT 2166 WD 4.  
M DOMAIN 1545 BEACH.  
M CONFLICT 1893 F->N (IN REF. 5).  
M SEQUENCE 2167 AA; 250870 MW; ED7D607ADA982CD CRC64;  
Query Match 4.7%; Score 83.5; DB 1; Length 2167;  
Best Local Similarity 18.1%; Pred. No. 77;  
Matches 48; Conservative 52; Mismatches 96; Indels 69; Gaps 9;  
Jy 115 FMCNGACAFEGDDVHWKENG-TLVQVATISGNMF--NQAKWVQKDN----- 162  
Jb 1170 FFMNTQVRFNITNTNITFKNNNSPISVRIKXQVYWKARBEYVQNNKKCLILFRK 1229  
Jy 163 --TGIYH-----TNVVASPKGAETWFDSDYCKEYLR-----TFNKLAERGA 205  
Jb 1230 DNTSIDFEXIKKISRYTYNLTKTRENNVAFYRNINILFIHKLTLIQSNNPNSCKWSS 1289

Qy 206 EFKNIETVTRIFLYSGEPTYL-----GNETSVPFGTNGKTLGLAIKRFYPPKPLP 258  
Db 1290 DAEDFDGKRRLLPLAMEPKXEPLINEDANDDTITG--GNQGRSSGILSYEFIEHME 1347  
Qy 259 TKE-----FLSLQIPAVVHKQFYLPNFEYF 289  
Db 1348 TLESFVGDINENRRILRLKXNDISIAITWNCSLIGLEIKGILIHGSNLYAFVSDYF 1407  
Qy 290 LPMKFPFIKITVEELPLPIRNTLS 314  
Db 1408 SLIEDKKILKLS--EVSQESRDWTVS 1430  
RESULT 11  
YCOL\_ARCFU  
ID YCOL\_ARCFU STANDARD; PRT; 228 AA.  
AC 029067;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF1201.  
GN AF1201.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxId=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.A., Gwin M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirnesh E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.,  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.",  
RL Nature 390:364-370(1997).  
C -----  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
C -----  
C EMBL; AB001021; AAB90048.1; -  
DR PIR; H69399; H69399.  
DR TIGR; AF1201; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 228 AA; 27209 MW; 3B29C4F59A26B68 CRC64;  
Query Match 4.7%; Score 83; DB 1; Length 228;  
Best Local Similarity 24.6%; Pred. No. 4.9;  
Matches 29; Conservative 24; Mismatches 55; Indels 10; Gaps 5;  
Qy 169 TNVAVASPKGAETWFDSDYCKEYLR-TFNKLAERGAETVTRIFLYSGEPTYL- 226  
Db 5 SMDRSGEQ--FHPTSCVDYILRIEKKFLFELCKIKMKKTKVVRGTYVQDAKX 60  
Qy 227 -LGNETSVFPGTNGKTLGLAIKRFYPPKPLPKEFL--SLQIFDAVIVHKQFYL 281  
Db 61 LIRLITELYNQSN-VLDQAIRLYDVSKYKSKADRIILREKLINPENILISKNEF 117

```
ESULT 12
122 HUMAN STANDARD; PRT; 889 AA.
09BZ06; OHCHW1;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Purified alpha-mannosidase Clorf22 (EC 3.2.1.-).
C1ORF22.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=21218927; PubMed=11318611;
Sood R., Bonner T.I., Makalowska I., Stephan D.A., Robbins C.M.,
Comors T.D., Morgensesser S.D., Su K., Faruqe M.U., Pinkett H.,
Graham C., Baxevanis A.D., Klinger K.W., Landes G.M., Trent J.M.,
Carpenter J.D.,
"Cloning and characterization of 13 novel transcripts and the human
RG88 gene from the 1q25 region encompassing the hereditary prostate
cancer (HPC1) locus."
Genomics 73:211-222(2001).
[2]
SEQUENCE OF 16-469 FROM N.A.
Cobley V.,
Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: Belongs to family 47 of glycosyl hydrolases.
-----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
-----
R EMBL; AF288393; MAGE6013.1; -.
R EMBL; AL096819; CAC03447.1; -.
R HSSP; P32906; 1D12.
R Genew; HGNC:16787; Clorf22.
R InterPro; IPR000886; ER_target_S.
R InterPro; IPR001382; Glyco_hydro_47.
R InterPro; IPR003137; PA.
R Pfam; PF01532; Glyco_hydro_47; 1.
R Pfam; PF02225; PA; 1.
R PRINTS; PR00747; GLYHDLASE47.
R ProDom; PD003239; Glyco_hydro_47; 1.
R PROSITE; PS00014; ER_TARGET; 1.
R Hypothetical protein; Hydrolase; Glycosidase; Glycoprotein;
R Endoplasmic reticulum.
W SITE 886 889 PREVENT SECRETION FROM ER (POTENTIAL).
T CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 771 771 N-LINKED (GLCNAC. . .) (POTENTIAL).
T CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
Q SEQUENCE 889 AA; 100304 MW; 21354A62C5901666 CRC64;

Query Match 4.7%; Score 83; DB 1; Length 889;
Best Local Similarity 20.6%; Pred. No. 27;
Matches 65; Conservative 41; Mismatches 108; Indels 102; Gaps 17;

44 MEGDDIEVRGAPWEEFYGDLLG-HLKIIMHDAIGFSTLTGKNYTWENY-ELFQDL 99
b 93 LNDNVVVSFEETIRY-IGLILGHSLALM-LKKGGEYMWQWINDLLQM 139
y 100 G-----NCTFPHLRPEMDAPFWCNOGACFPEGIDVHWKENG 137
b 140 AKOLGYKLLPAFNTTSGLPYPRINLKFKIRKPEARAGTETDTCTAC-----AG 187
```

```
QY 138 TLV-QVATIS-----GNMNEQYAK-----WVKQDNETGIYYETWNVKAS-----PEKGAET 182
Db 188 TLIEFALSLFTGATIFEEYARALDPLMEKFORSSNLVGTITINIRHGDWVRKSGVGA 247
QY 183 WEDSYDSKFLVLRPNKLAEGAEFKNIETNYTRIFLYSGEPTIYGNTSVGPFQNTL 242
Db 248 GIDSY-VEYLLKAYVLGD-DSFLERENTHYDAIMRYISQPPPL-----L 290
QY 243 GLATKRFYPPKPHLPTEKEFLSLQLF-----DAVIYHKQYLFYFPEWFLP 291
Db 291 DVHI-----HKPMINATWMDALALAFPPGQVLKGDIRPALETHEMLYQYK-KANFLP 343
QY 292 MKPFP-IKITYEELPL 306
Db 344 EAFITDFRVHNAQHPL 359

RESULT 13
HGFB_HAEBIN STANDARD; PRT; 999 AA.
AC 087256;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein B precursor.
GN HGFB.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Hi689 / Serotype B;
RA MEDLINE=98427137; PubMed=9746572;
RX Ren Z., Jin H., Morton D.J., Stull T.L.;
RT "hgfb, a gene encoding a second Haemophilus influenzae hemoglobin- and
RT hemoglobin-haptoglobin-binding protein."
RT Infect. Immun. 66:4733-4741(1998).
[2]
SEQUENCE FROM N.A.
RP STRAIN=Ela / serotype B;
RC Morton D.J., Stull T.L.;
RA "Conservation of hemoglobin/hemoglobin-haptoglobin binding proteins in
RT Haemophilus influenzae."
RU Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPARIING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING OF THE
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF022910; AAC60790.1; -.
CC DR EMBL; AF259266; AAK51630.1; -.
CC DR InterPro; IPR006970; PT.
CC DR InterPro; IPR00531; Tonb_boxC.
CC Pfam; PF04886; PT; 1.
```

Pfam: PF00593; TonB\_dep Rec; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 CW Outer membrane; Transport; TonB box; Multigene family; signal;  
 CW Receptor; Repeat.  
 FT SIGNAL 1 24  
 FT CHAIN 25 999  
 FT DOMAIN 26 57  
 FT REPEAT 26 29  
 FT REPEAT 30 33  
 FT REPEAT 34 37  
 FT REPEAT 38 41  
 FT REPEAT 42 45  
 FT REPEAT 46 49  
 FT REPEAT 50 53  
 FT REPEAT 54 57  
 FT SITE 66 73  
 FT SITE 982 999  
 FT VARIANT 46 57  
 FT VARIANT 85 101  
 FT VARIANT 101 101  
 FT VARIANT 108 108  
 FT VARIANT 121 122  
 FT VARIANT 151 151  
 FT VARIANT 304 304  
 FT VARIANT 560 560  
 FT VARIANT 633 633  
 FT VARIANT 658 658  
 FT VARIANT 689 689  
 FT VARIANT 755 755  
 FT VARIANT 782 782  
 FT SEQUENCE 999 AA; 114435 MW; 58F631FA5D2685B0 CRC64;  
 Query Match 4.7%; Score 83; DB 1; Length 999;  
 Best Local Similarity 18.7%; Pred. No. 32; Index 160; Gaps 20;  
 Matches 71; Conservative 44; Mismatches 105; Indels 160; Gaps 20;  
 16 KRDFRPPDPYCOAKYTC-----PTGSPDPMEGDDIEVRLQAPWVEFKYGDILGH 70  
 370 QRIKTRARTEYCDAGVRHCEGTDPNG--LKYTDG---KTRRROSELOFKKNTAH 423  
 71 -----LKMIDAIQFSTLTGKQYTWEMWYLFQLGCTPFLHPREMDAPF 115  
 424 SNGRYVDFFKPIDTDKKVIED-----KLTLKPNPDWYDC-SIFNCE----- 464  
 116 WCGAGACFEFGI---DQVHKENGTLVQATISGMFNQMAK----- 155  
 465 --NNAKIVFEGNYSYGYDGKMKR--VDLKITLNGKKFAKITOPNKKISILPSSPGYL 520  
 156 ---WVKQDNTEGLY-----YETWNVKASPEKGAETWFDSDSKFVLTFTFNKLAEP 203  
 521 ERLMQERDLDLTNQNLNLDLTQPKTWHEHNLQYGS---SYNTA--MKRWVNR----- 569  
 204 GAEFKNIETNYTLFLYSGEPTLYGNETSYFGPTGNKTLGLAKRFY-----YPRK 254  
 570 -----AGNDASDLYQWMAAPITLGY---NRYDQRYCATAYSWN 603  
 255 PHL-----PTKEFLSLILOFDVAIVHKQSYLYNF---EYWFLPKKPFRIKITYAE--- 303  
 604 ANLCPRVDPEFSYLLP-----KTTGKSYVLPDNLVITDYSFDLYGRYDNIHYQPKYK 657  
 304 -----IPLP 307  
 658 RGIIPKLPDDIYKGLFPLP 677  
 RESULT 14  
 PDGK HAEIN STANDARD; PRT, 1028 AA.  
 AC P46448;  
 JT 01-NOV-1995 (Rel. 32, Created)  
 JT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Formate dehydrogenase major subunit (EC 1.2.1.2) (Formate  
 dehydrogenase alpha subunit) (FDH alpha subunit).  
 GN PDGK OR H10006  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Heddlow E., Cotton M.D.,  
 RA Ullrich L.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd".  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: ALLOWS TO USE FORMATE AS MAJOR ELECTRON DONOR DURING  
 CC ANAEROBIC RESPIRATION. SUBUNIT ALPHA POSSIBLY FORMS THE ACTIVE  
 CC SITE.  
 CC -1- CATALYTIC ACTIVITY: Formate + NAD(+) = CO(2) + NADH.  
 CC -1- COFACTOR: Molybdenum (molybdopterin) and selenocysteine. The  
 CC active-site selenocysteine is encoded by the opal codon, UGA. May  
 CC bind a 4Fe-4S cluster.  
 CC -1- SUBUNIT: FORMATE DEHYDROGENASE IS A MEMBRANE-BOUND COMPLEX, FORMED  
 CC BY SUBUNITS ALPHA, BETA AND GAMMA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).  
 CC -1- SIMILARITY: ORTHOLOG OF BOTH E.COLI FDG AND PDGK.  
 CC -1- SIMILARITY: Belongs to the prokaryotic molybdopterin-containing  
 CC oxidoreductase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL, U2866; -, NOT\_ANNOTATED\_CDS.  
 DR TIGR; H10006; -;  
 DR InterPro; IPR009010; Asp. decarb. fold.  
 DR InterPro; IPR006443; Formate-dh-alph.  
 DR InterPro; IPR006657; Mol. dinuc. bind.  
 DR InterPro; IPR006963; Molybdop. Fe4S4.  
 DR InterPro; IPR006656; Molybdopterin.  
 DR InterPro; IPR006655; Prok. Moxored.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF04879; Molybdop. Fe4S4; 1.  
 DR Pfam; PF00384; molybdopterin; 1.  
 DR Pfam; PF01568; Molybdop. binding; 1.  
 DR TIGRPFAM; TIGR01553; formate-dh-alph; 1.  
 DR TIGRPFAM; TIGR01409; Tat. signal seq; 1.  
 DR PROSITE; PS00551; Molybdopterin PROK\_1; 1.  
 DR PROSITE; PS00490; Molybdopterin PROK\_2; FALSE\_NEG.  
 DR PROSITE; PS00932; Molybdopterin PROK\_3; 1.  
 DR Oxidoreductase; Molybdenum; Selenocysteine; Selenium; NAD;  
 KM Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 50 50  
 FT METAL 53 53  
 FT METAL 57 57  
 FT METAL 100 100  
 FT SE CYS 204 204  
 FT SEQUENCE 1028 AA; 115403 MW; 295F185BB020EDP9 CRC64;  
 Query Match 4.7%; Score 83; DB 1; Length 1028;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## 4 protein - protein search, using sw model

on on: June 8, 2004, 13:59:08 ; Search time 58.3722 Seconds

(without alignment)  
1708.068 Million cell updates/sec

Title: US-10-010-050a-2\_COPY\_31\_346

Sequence: 1 RVSGSPSRHMPVPRKDF.....IKIYEELPIRKMTLSGL 316

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaea:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1423	80.9	336	11 Q8R152	Q8R152 mus musculu
2	1418	80.7	322	11 Q8C054	Q8C054 mus musculu
3	235	13.4	180	5 Q86157	Q86157 dictyosteli
4	202	11.5	378	5 Q86156	Q86156 dictyosteli
5	101	5.7	633	3 Q8N1H1	Q8N1H1 trichophyto
6	99.5	5.7	514	16 Q8A3V2	Q8A3V2 bacteroides
7	98	5.6	682	4 Q8NBV6	Q8NBV6 homo sapien
8	97.5	5.5	804	5 Q8IAU5	Q8IAU5 plasmodium
9	96	5.4	508	5 Q8IDS4	Q8IDS4 plasmodium
10	95.5	5.4	359	10 Q9ZV89	Q9ZV89 arabidopsis
11	95.5	5.4	836	4 Q13785	Q13785 homo sapien
12	95.5	5.4	5322	5 Q9VPL9	Q9VPL9 drosophila
13	94	5.3	362	5 Q9TYM2	Q9TYM2 caenorhabdi
14	94	5.3	1338	16 Q25330	Q25330 haemobacte
15	94	5.3	2313	8 Q8WHM9	Q8WHM9 psittacus nu
16	93	5.3	3118	5 O17575	O17575 caenorhabdi

17	92.5	5.3	564	5 Q8MQ12	Q8MQ12 drosophila
18	92.5	5.3	569	5 Q9VEY2	Q9VEY2 drosophila
19	92	5.2	502	16 Q65561	Q65561 aquifex ae
20	91	5.2	291	5 Q01702	Q01702 caenorhabdi
21	91	5.2	508	10 Q9XEC4	Q9XEC4 arabidopsis
22	91	5.2	607	4 Q9BVZ9	Q9BVZ9 homo sapien
23	91	5.2	742	4 Q94841	Q94841 homo sapien
24	91	5.2	742	4 Q72316	Q72316 homo sapien
25	90.5	5.1	850	2 Q8VJF6	Q8VJF6 azaracus ev
26	90.5	5.1	1846	5 Q81421	Q81421 plasmodium
27	90	5.1	1409	3 Q7Z918	Q7Z918 coprinus ci
28	89.5	5.1	1512	2 Q9GXT5	Q9GXT5 streptococc
29	89	5.1	567	5 Q9GXT5	Q9GXT5 drosophila
30	89	5.1	570	5 Q8M1N6	Q8M1N6 drosophila
31	89	5.1	584	5 Q8MT35	Q8MT35 drosophila
32	89	5.1	613	5 Q9W122	Q9W122 drosophila
33	89	5.1	1353	5 Q9XW65	Q9XW65 caenorhabdi
34	88.5	5.0	465	10 Q9ZS50	Q9ZS50 ipomoea bat
35	88.5	5.0	540	2 Q9AIP9	Q9AIP9 methylococc
36	88.5	5.0	1462	4 Q86U07	Q86U07 homo sapien
37	88	5.0	742	6 Q7Z9M6	Q7Z9M6 trichoderma
38	88	5.0	742	6 Q29397	Q29397 bos taurus
39	88	5.0	2107	12 Q91NQ2	Q91NQ2 tauro synter
40	88	5.0	2713	5 Q8MMW0	Q8MMW0 dictyosteli
41	88	5.0	3194	16 Q9Z1M3	Q9Z1M3 helicobacte
42	87.5	5.0	755	12 Q56271	Q56271 human herpe
43	87.5	5.0	809	5 Q9N3K6	Q9N3K6 caenorhabdi
44	87.5	5.0	1162	3 Q9UTH0	Q9UTH0 schizosacch
45	87	4.9	251	5 Q18207	Q18207 caenorhabdi

## ALIGNMENTS

```
RESULT 1
Q8R152 PRELIMINARY; PRT; 336 AA.
ID Q8R152;
AC Q8R152;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Similar to ceroid-lipofuscinosis, neuronal 5 (Fragment).
GN C1N5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025487; AAH25487.1; -.
DR MGD; MGI:2442253; C1N5.
FT NON TER 1
SQ SEQUENCE 336 AA; 38911 MW; D50D74C45483BE1B CRC64;
```

Query Match 80.9%; Score 1423; DB 11; Length 336;  
Best Local Similarity 79.4%; Pred. No. 3.1e-116;  
Matches 246; Conservative 30; Mismatches 34; Indels 0; Gaps 0;

QY	7	SRHMPVPRKDFPPKPPDYCOAKYTCPTSPPIPVNMGDDIEVFRLOAPVWERKYD	66
QY	1		66
DB	27	SGQRPVPRKSRFPKDPYCOAKYTCPTSPPIPVNMGDDIEVFRLOAPVWERKYD	86
QY	67	ILGHKIMEDALIGFSTLTGKNYTWYELFOLGNCPEPHLRPEMDAPWCOGAACPE	126
QY	1		126
DB	87	ILGHKIMEDALIGFSTLTGKNYTWYELFOLGNCPEPHLRPEMDAPWCOGAACPE	146
QY	127	GIDVYHKNENGLVQVATISGNENOMAKMYKODNETGIVYETWVKSPEKGAETWFS	186
QY	1		186
DB	147	GIDVYHKNENGLVQVATISGNENOMAKMYKODNETGIVYETWVKSPEKGAETWFS	206
QY	187	YDSKFLVLTFFNKLAFGAEFNKINETNYRIFLYSGEPYLGNETSVFGTGNKTLGLAI	246
QY	1		246



207	YDGSNFVARTYKGLAEFETFEKFKIETNYTKTFLKSSGPYLGNPTSIFGKGNKTKLALAI	126
247	KRPYYPRKPHLPTEKEFLLSLQIFDAVIYHKQFYLFFNFEYEWFLPMKFPFIKITYEIPL	306
267	KKFYGPFRPYLSTKDFLNMFLKIFDVIYIHRQFYLFFNFEYEWFLPMKPPPVKITYEETPL	326
307	PIRNKTLISGL 316	
327	PTRHHTFTDL 336	
RESULT 2		
8C054		
Q8C054	PRELIMINARY; PRT; 322 AA.	
01-MAR-2003	(TREMBLrel. 23, Created)	
01-MAR-2003	(TREMBLrel. 23, Last sequence update)	
01-JUN-2003	(TREMBLrel. 24, Last annotation update)	
Similar to CEROID-LIPOFUSCINOSIS neuronal protein 5 (Fragment).		
Mus musculus (Mouse).		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.		
NCBI_TaxId=10090;		
SEQUENCE FROM N.A.		
STRAIN=C57BL/6J; TISSUE=Olfactory brain;		
MEDLINE=22354683; PubMed=12466851;		
The FANTOM Consortium,		
the RIKEN Genome Exploration Research Group Phase I & II Team;		
"Analysis of the mouse transcriptome based on functional annotation of		
60,770 full-length cDNAs."		
Nature 420:563-573(2002).		
EMBL; AK032293; BAC27797.1; -		
MGD; MGI:2442253; Cln5.		
NON TER		
1		
Q	SEQUENCE 322 AA; 37338 MW; 04FE01B05B0DED8A CRC64;	
Query Match 80.7%; Score 1418; DB 11; Length 322;		
Best local similarity 79.0%; Pred. NO. 8e-116;		
Matches 245; Conservative 30; Mismatches 35; Indels 0; Gaps 0;		
Y	7 SRHMPVPEYKFRDEPRKPYCAQKTYFCCTGSPIDVMEGDDIEYFRLQAEVWEFKYGD	66
Y	13 SGQWNPVYKHFSPRPKTDYCAQKTYFCCTGSPIDVMEGDDIEYFRLQAEVWEFKYGD	72
Y	67 LIGHKLYMDAIGRSTLTGKNTMYEYLFQGNCTFPHLRPMDAPWCNOGAACFE	126
Y	73 LIGHKLYMDAIGRSTLTGKNTMYEYLFQGNCTFPHLRPMDAPWCNOGAACFE	132
Y	127 GIDVHWKENQILQVATTISGNMKNQAKVKKDNETGIYETFWYVWKAPEEGAEFTWDS	186
Y	133 GIDVHWKENQILQVATTISGNMKNQAKVKKDNETGIYETFWYVWKAPEEGAEFTWDS	192
Y	187 YDGSNFVARTYKGLAEFETFEKFKIETNYTKTFLKSSGPYLGNPTSIFGKGNKTKLALAI	246
Y	193 YDGSNFVARTYKGLAEFETFEKFKIETNYTKTFLKSSGPYLGNPTSIFGKGNKTKLALAI	252
Y	247 KRPYYPRKPHLPTEKEFLLSLQIFDAVIYHKQFYLFFNFEYEWFLPMKFPFIKITYEIPL	306
Y	253 KRPYYPRKPHLPTEKEFLLSLQIFDAVIYHKQFYLFFNFEYEWFLPMKFPFIKITYEIPL	312
Y	307 PIRNKTLSGL 316	
Y	313 PTRHHTFTDL 322	
RESULT 3		
86157		
086157	PRELIMINARY; PRT; 180 AA.	
01-JUN-2003	(TrEMBLrel. 24, Created)	

```

DR 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein.
OC Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafrański K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116362; AAC51609.1; -.
KW Hypothetical protein.
SQ SEQUENCE 180 AA; 20631 MW; 2DEB66F4FBCT1AC CRC64;

Query Match 13.4%; Score 235; DB 5; Length 180;
Best Local Similarity 35.8%; Pred. No. 1,2e-12;
Matches 54; Conservative 25; Mismatches 60; Indels 12; Gaps 6

QY 22 PKPDP-YCOAQK--TFCPGTGSPFP--VMEGDDILEVFLQAPWMEFFKGDILGHKIMED 76
DB 23 PDNDPELCOQOQIQREDPCQ-TPVPMGTGFNDDELBYVYMQAPVEAVGNGFFGKGGYHS 81
QY 77 AIGRSTLTGKNTYMEWVEYELFQLCNCTEPHL--RPMDAPFMCNCGAACFFEGIDDVHMK 134
DB 82 AIGGYDILTGINTYAEADAPFEVNGNLTPIINVDGKDELIMCAGILCTVPYINMETYMD 141
QY 135 EN---GLTVQVATISGNMVENQAKWVKQDN 161
DB 142 KNIVTASKTYMTGINTGTLNQYIEVMQYVN 172

RESULT 4
QID Q86JG6 PRELIMINARY; PRT; 378 AA.
AC Q86JG6;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DS Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafrański K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116362; AAC51609.1; -.
KW Hypothetical protein.
SQ SEQUENCE 378 AA; 42306 MW; 2CF33D60851FD036 CRC64;

Query Match 11.5%; Score 202; DB 5; Length 378;
Best Local Similarity 23.4%; Pred. No. 2,4e-09;
Matches 65; Conservative 53; Mismatches 130; Indels 30; Gaps 8;

```



[illegible]

RESULT 7		
Q8NB36		
ID	Q8NB36	PRELIMINARY; PRT; 682 AA.
AC	Q8NB36;	
DT	01-OCT-2002 (TREMblrel. 22, Created)	
DT	01-OCT-2002 (TREMblrel. 22, Last sequence update)	
DT	01-OCT-2003 (TREMblrel. 25, Last annotation update)	
DE	Hypothetical protein NTFRP2002760.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,	
RA	Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,	

RA Nagahari K., Sugano S., Isogai T.;  
 RT "HRI human cDNA sequencing project";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AK075480; BAC11645.1; -  
 DR GO: GO:0016021; C:Integral to membrane; IEA.  
 DR GO: GO:0005215; F:Transporter activity; IEA.  
 DR GO: GO:0006810; P:Transporter; IEA.  
 DR InterPro: IPR007114; MFS.  
 DR InterPro: IPR005828; Sub transporter.  
 DR InterPro: IPR005828; Sug\_transporter.  
 DR InterPro: IPR005988; SV2.  
 DR Pfam: PF00083; Sugar\_tr; 1.  
 DR TIGRFAMs: TIGR01299; synaptc\_sv2; 1.  
 DR PROSITE: PS00850; MFS; 1.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 DR Hypothetical protein, Transmembrane.  
 KM SEQUENCE 682 AA; 7660 MW; DIA76C115424CCTF CRC64;  
 SQ

Query Match 5.6%; Score 98; DB 4; Length 682;  
 Best Local Similarity 22.1%; Pred. No. 6.6;  
 Matches 62; Conservative 40; Mismatches 104; Indels 74; Gaps 16;

QY 86 GKNYTWMEYELFQGNCTPEH-----LRPMDAP-FKCNQG---AACFEGIDVHWK 134  
 DB 325 GSAYQFHSRWYFVL-VCAFPSPVFAIGALTTPESPSPFLENGKDEAMVILKQYHDTNMR 383  
 QY 135 ENGLTVQVATISG-NMFQMAKWKVKODNETGIYETWNVKASPEKGAETWDSYDCKEV 193  
 DB 384 AKGPERFVSVTHHTIHQEGHLEIDSDGTWYGRWVRL-SLGGVWGNFLSC----- 438  
 QY 194 LRTFNKLAFFGAFFKNIEFN---YTRIFLYSGEPTL-----GNESVFP-- 234  
 DB 439 -----FGPEYRITLMMWGVFTMSFSYGLTWPEPMIRHLQAVDASRTKXVPPG 489  
 QY 235 ---GPTGNKTLGLAIKRFYFEPK-----HLPTKEFLSLIQ--IFDAVIYHKOY--- 280  
 DB 490 KEVHEVTEFTLENDIHRGQYFNDKFLGLRKSASFEDSLFECEYFEDVTSNNAFFKNC 549  
 QY 281 -----LEFN---FEYWFLPMKFFIKITY---EEIPLPI 308  
 DB 550 TEINTVFINTDLFEYKFVNSR--LINSTPLHNKEDCPIDV 587

## RESULT 8

Q81AUS PRELIMINARY; PRT; 804 AA.

AC Q81AUS; 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)  
 DE Cullin-like protein, putative.  
 GN PF08\_0094.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL844507; CAD51265.1; -  
 DR GO: GO:0007049; P:cell cycle; IEA.  
 DR InterPro: IPR001373; Cullin.  
 DR Pfam: PF00888; Cullin; 1.  
 DR SMART: SM00182; CULLIN; 1.  
 DR PROSITE: PS50069; CULLIN\_2; 1.  
 SQ SEQUENCE 804 AA; 97162 MW; 6B6E916BA95AD2C CRC64;

Query Match 5.5%; Score 97.5; DB 5; Length 804;  
 Best Local Similarity 19.8%; Pred. No. 9;  
 Matches 57; Conservative 51; Mismatches 115; Indels 65; Gaps 14;

QY 23 KEDPYQQAQKTTCPFGSP-----PVMGDDIDIEVFR-QAPWMEKY-----GD 66  
 DB 57 KKNPFCYSKEVYRKXGSELSMTYTDKIKPLKNSDELNKTLLIDAM-FKYSFYYKMMNK 115  
 QY 67 ILGH;KIMEDALGFRSTLTKKNTMWEYELFQGNCTPEHLPREMDAPWCNQGAACFE 126  
 DB 116 FLRYLDRIY-VEYNSSLCSAVTKNIFK-----TLFVELR-----E 151  
 QY 127 GIDDVHWK-ENGLTVQVATISGMENQMAKWKQ-DNEG-IYYE-----TWNVKAS 176  
 DB 152 DIKNITVEIYNLRLGEEIDQKELFCNIVEIYKELDNESNKKYEHDIKKIVENNNFY 211  
 QY 177 EKCAETWDSYDCKEVLTENKLAFFGAFFKNIEFN-----YTRIFLYSGEPTL 226  
 DB 212 KKKAEEMINDYPPDDYIIETENAIKEYEKNSLNINDTCEKVTNIYKILLIEKNTL 271  
 QY 227 LGNETSVFGPTGNKTLGLAIKRFYFEPKPHLPKFEFLSLIQIFDAVI 274  
 DB 272 IDNKNNTFHLKNNNLS-SLRRTYILFS-YFP-EALTGKILIGEVY 315

## RESULT 9

Q81DS4 PRELIMINARY; PRT; 508 AA.

AC Q81DS4; 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Hypothetical protein.  
 GN PF13\_0225.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Harris B., Denard N., Clark L., Line A., Barron A., Corton C.,  
 RA Berriman M., Pain A., Hall N., Actin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL844509; CAD52545.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 508 AA; 58842 MW; F563E34C1A9F25D2 CRC64;

Query Match 5.5%; Score 96; DB 5; Length 508;  
 Best Local Similarity 20.0%; Pred. No. 6.7;  
 Matches 52; Conservative 33; Mismatches 95; Indels 80; Gaps 11;

QY 126 EGIDVHWKENGTLVQVATISGMENQMAKWKVKODNETGIYEE-----TWNVKASPEKG 179  
 DB 121 DINKNNNPHNNNNYNNVYTLDNKPMKSKSMNDPNSNITNMINDGIMNKEVPSNI 180  
 QY 180 AETWDSYDCKEVLRTEFN-----KLAFFGAFFKNIEFN-YTRIFLYS 221  
 DB 181 EN-YDYVDIPNPQVNNQOLYSANNIYVPLPYGRVYKIOENDSINGVNTNEYIKGHIV 238  
 QY 222 GE-----PYL-----GNESVFGPTGNKTL 242  
 DB 239 GRKPFNVLAGCOLPSLCTPSYTTSSNNNNNNNNNNNNNNNNNNNNIILGPPENKTL 298  
 QY 243 GLAI-----KRYVEFKPHLPKFEFLSLIQIFDAVIYHKOYLFNFEYFLPMKFFP 296  
 DB 299 NMEYIINDONKEKITYIY-----LNKGNNELTNDYNNWSTNK-FISPNYEYLMTNPSPY 351  
 QY 297 IKITY-EEIPLPIRNK-TLS 314  
 DB 352 YMLNINKEKLDLFFKKTLLS 371

## RESULT 10

Q9ZV89 PRELIMINARY; PRT; 359 AA.

ID Q9ZV89  
 AC Q9ZV89; 01-MAY-1999 (TrEMBLrel. 10, Created)

```

1 01-MAY-1999 (TrEMBLrel. 10, last sequence update)
2 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
3 F9K20.25.
4 F9K20.25.
5 Arabidopsis thaliana (Mouse-ear cress).
6 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
7 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
8 eurosids II; Brassicales; Brassicaceae; Arabidopsis.
9 NCB1_TaxID=3702;
10 [1]
11 SEQUENCE FROM N.A.
12 STRAIN=cv. Columbia;
13 Vystotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Li J., Liu S.,
14 Kremetskaya I., Iurov J., Araujo R., Buehler E., Conway A.B.,
15 Dewar K., Peng J., Kim G., Li Y., Shinn P., Davis R.W., Ecker J.R.,
16 Federspiel N.A., Theologis A.,
17 "Arabidopsis thaliana chromosome 1 BAC F9K20 sequence."
18 Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
19 (2)
20 SEQUENCE FROM N.A.
21 STRAIN=cv. Columbia;
22 Theologis A.;
23 Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
24 EMBL; AC005679; AAC83039.1; -.
25 PIR; A96816; A96816.
26 InterPro; IPR004253; DUF231.
27 Pfam; PF03005; DUF231; 1.
28 PIRam; PF03005; DUF231; 1.
29 SEQUENCE 359 AA; 40568 MW; A9887DD29786278B CRC64;
30
31 Query Match 5.4%; Score 95.5; DB 10; Length 359;
32 Best Local Similarity 21.5%; Pred. No. 4.7;
33 Matches 53; Conservative 39; Mismatches 97; Indels 57; Gaps 12;
34
35 32 YTEGCTGPIPMEDDDIEVRLQAPVMEFKYGDLLHKLMDAIGFSTLTGKATM 91
36 80 YRMQPTGDIIPENGRDPLTRKKGKILF---VGDLSLN-----N 116
37 92 EWEYELFQGNCTFPHLRPEMDAPFWCNOGAACFF---EGIDVHKENGTLVQVAT--- 144
38 117 MWVSL---SCMLHAAYVNAKTFPQANKGLSTFTPEGI--SYNFKMGFLVDVSDGTR 171
39 145 ---ISGNFNOAKAVKQDNETGIYETWNVKASPEKGAETWFSYDCKFVLRTFNKLA 201
40 172 GLIKLDSISRGNOMLISDVAIFNTEFWW---SHTGRAKTW-DYFGDGIKIVEMNME 226
41 202 EFGAEF-----KLEETNTRIELFVSGEPTYL-GNETSVFGPIGAKTLG--LAIKRFY 250
42 227 AKKIALTWKMKIDHNIDPSKTRVFGVSVPHLNGS---WGKRGKTCIGETVFGQPS 283
43
44 251 YPFKEH 256
45 284 YPGRPN 289
46
47 SUPT 11
48 3785 PRELIMINARY; PRT; 836 AA.
49 013785.
50 01-NOV-1996 (TrEMBLrel. 01, Created)
51 01-NOV-1996 (TrEMBLrel. 01, last sequence update)
52 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
53 APOB protein (Fragment).
54 APOB.
55 Homo sapiens (Human).
56 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
57 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
58 NCB1_TaxID=9606;
59 [1]
60 SEQUENCE FROM N.A.
61 MEDLINE=86042646; PubMed=2932736;
62 Wei C.F., Chen S.H., Yang C.Y., Marcel Y.L., Milne R.W., Li W.H.,
63 Sparrow D.T., Goto A.M., Jr., Chan L.,
64 "Molecular cloning and expression of partial cDNAs and deduced amino

```

```

RT acid sequence of a carboxyl-terminal fragment of human apolipoprotein
RT B-100 *;
RL Proc. Natl. Acad. Sci. U.S.A. 82:7265-7269 (1985).
RL EMBL; M12413; AAA51742.1; -.
DR PIR; A27850; LPHD.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000531; T0NB_BoxC.
DR PROSITE; PS00430; T0NB_DEPENDENT_REC_1; 1.
FT NON TER 1
SQ SEQUENCE 836 AA; 96317 MW; 5298C7EE21AB140 CRC64;
37
38 Query Match 5.4%; Score 95.5; DB 4; Length 836;
39 Best Local Similarity 20.5%; Pred. No. 14;
40 Matches 61; Conservative 39; Mismatches 125; Indels 73; Gaps 12;
41
42 37 TGSPIPV-----MEGDDIEVRLQAPVMEFKYGDLLG---HLKIMDAIGFSTLTG 86
43 272 TSAPSPAVGTGMVDDEDDDS-----KNNFTYSPSSPDCKLTITKTELRVRESDE 324
44
45 87 KNYTWWEYELFQGNCTFPHLRPEMDAPFWCNOGAACFFEGIDVHKENG-TLVQVAT 145
46 325 TQIKNWEBAASGLT--SLKDNVP-----KATGVLYDVNKKVHMTGTLAEVSK 376
47
48 146 SGNNFNOAKAVKQDNETGIYETWNVKASPEKGAETWFSYDCKFVLRTFNKLAERGA 205
49 377 LRRNIQDHAEWVYQ---GAIRETDIDDERFOKSGASTGTGYQ----- 415
50
51 206 EFKNIETN-YRIFLYSGEPTYLGNETSVPFGTNGKTLGLAIK-----RF 249
52 416 EFKDKAQNLYBELITQSGQASFGIKDNVFDGLVAVTQFFMKKHLIDSLIDFNFRF 475
53
54 250 YPFKEHLPTEFFLSLQIFDAVIVHKQFY-----LFYNEEYVLPKPPFIK 298
55 476 QPFGKPIGYTREELCTWP--IREVGTVLSQYVSKVHNGSELLFSYQDVLITLPFLRK 532
56
57 RESULT 12
58 09VPL9 PRELIMINARY; PRT; 5322 AA.
59 ID 09VPL9; Q9N164;
60 AC 09VPL9; Q9N164;
61 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
62 DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)
63 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
64 DE CG3696 protein (KISMET-L long isoform).
65 GN KIS OR CG3660 OR CG3696 OR CG18326.
66 OS Drosophila melanogaster (Fruit fly).
67 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
68 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
69 CC Ephydroidea; Drosophilidae; Drosophila.
70 NC NCB1_TaxID=7227;
71 [1]
72 SEQUENCE FROM N.A.
73 RC STRAIN=Berkley;
74 RC MEDLINE=20196006; PubMed=10731132;
75 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
76 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
77 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
78 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
79 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
80 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
81 RA Abail J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
82 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
83 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
84 RA Borokova D., Botchan M.R., Bouck J., Brockett P., Broctler P.,
85 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
86 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
87 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
88 RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
89 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiza S., Fleischmann W.,
90 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kamot I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*";  
 Science 287:2185-2195(2000).  
 [12]  
 SEQUENCE FROM N.A.  
 Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 Evans C.A., Goebye U.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,  
 Carlson J.W., Center A., Change M., Davenport L.B., Dietz S.M.,  
 Dodson K., Doreste V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 Ferreira S., Frise E., Galle R.F., Gang N.S., George R.A.,  
 Gonzalez M., Hunk J., Hoskins R.A., Hostin D., Howland T.J.,  
 Ileguam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,  
 Pacleb J., Pargass V., Park S., Patel S., Pfeiffer B.,  
 Pounanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 Stapleton M., Strong R., Svrtkars R., Tector C., Tyler D.,  
 Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 "Sequencing of *Drosophila melanogaster* genome";  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [13]  
 SEQUENCE FROM N.A.  
 Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 Tupy J.L., Bergman C., Bertman B., Carlson J.W., Celniker S.E.,  
 Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 "Annotation of *Drosophila melanogaster* genome";  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [14]  
 SEQUENCE FROM N.A.  
 Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [15]  
 SEQUENCE FROM N.A.  
 Flybase;  
 Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 [16]  
 SEQUENCE FROM N.A.  
 Therrien M., Morrison D.K., Wong A.M., Rubin G.M.;  
 "A Genetic Screen for Modifiers of a KSR-Dependent Rough Eye Phenotype  
 in *Drosophila*";  
 Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 EMBL; ABO03590; AAF51527.3; -  
 EMBL; AF215703; AAF43004.1; -  
 HSSP; P23197; IAP0.  
 Flybase; Fggn0001309; Kis.  
 GO; GO:0000785; C:chromatin; IEA.  
 GO; GO:0005634; C:nucleus; IEA.  
 GO; GO:0005524; C:ATP binding; IEA.  
 GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 GO; GO:0003682; F:chromatin binding; IEA.

DR GO:0016787; F:hydrolase activity; IEA.  
 DR GO:0006333; P:chromatin assembly/disassembly; IEA.  
 DR InterPro: IPR000953; Chromo.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR000330; SNF2\_N.  
 DR Pfam: PF00385; Chromo; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS50013; CHROMO\_2; 1.  
 KM ATP-binding; Helicase; Hydrolase.  
 SQ SEQUENCE 5322 AA; 573615 MW; C9608375FA71C211 CRC64;  
 Query Match 5.4%; Score 95.5; DB 5; Length 5322;  
 Best Local Similarity 20.1%; Pred. No. 1.6e+02;  
 Matches 71; Conservative 49; Mismatches 139; Indels 95; Gaps 17;  
 QY 1 RVSGIPSRHHMPVYKRPDPYCOAKYTFCTGSPIMMEDDDIEVRLQAPW 60  
 DB 1993 RNFKIPQSEW-----KSKRRPPELWKLKLT-----PYKGSNLRPYQLGLNW 2039  
 QY 61 EFKYGDILGHLXIMDAIGRSTLTKNTWMEYELFOLGNCFTPHLRPMADPFWNOG 120  
 DB 2040 -LKFSYNHNHCLDDEMLGKTIQSLTVHSGYEGIRG--PFLVIAPLSTIPWQR- 2095  
 QY 121 AACFFPGIDVHWKENGTLVQVATISGNFNOMAKVKQDNFTGIYFTWNVASPE-- 177  
 DB 2096 ---FEG-----WTDNANVVVHGSVT-----SKQIDYE--YYKTESGVLKEPIK 2137  
 QY 178 -KGAEFWPSYDSCKFVLRFTN-----KLAFFGAEFKNIE----- 211  
 DB 2138 FNVLTTFEMITVDYMDLKAFFWRLCVIDEARLKNRCKLIE-GIRQLNLRHVLISGT 2196  
 QY 212 ---TWYTRIF--LYSGEPYTLGNETSVEFGPTGNKTLGLAIKRFYFPKPHL----- 257  
 DB 2197 PLQNNISELFSILNLEPQSFSSQEFMSFSLTTEEVNKLQALKKMRLKLDYVE 2256  
 QY 258 ---PTKEFLSLLOLFDPAVIVHKOPYLFYNEFWFLPKFPPIK--ITYEIP 305  
 DB 2257 KSLAPKEETILIV---ELTNIQKKY-----RGLLEQNFSLKGTTSANIP 2300  
 RESULT 13  
 Q9TYM2 PRELIMINARY; PRT; 362 AA.  
 AC Q9TYM2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Y25CIA.7a protein.  
 GN Y25CIA.7 OR Y25CIA.7A.  
 OS Caenorhabditis elegans.  
 OC Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 Jones M., Kerhaw J., Kirsten T., Laister N., Latteille P.,  
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,  
 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 Thairon-Mieg J., Thomas K., Vaudin N., Vaughan K., Waterston R.,  
 Watson A., Weinstock L., Wilkison-Sproat V., Wohlman P.;  
 "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*

elegans";  
 Nature 368:32-38 (1994).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN=Bristol N2.  
 Kalkicki U., Smith A., Gibson A.;  
 "The sequence of *C. elegans* cmd Y25C1A.";  
 Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases  
 [3]  
 SEQUENCE FROM N.A.  
 STRAIN=Bristol N2.  
 Waterston R.;  
 Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases  
 EMBL: AF125459; AAD12638.1; -  
 PIR: T33904; T33904.  
 WormPep; Y25C1A.7a; C21474.  
 InterPro; IPRO06977; DUF649.  
 Pfam: PF04893; Ylip; 1  
 SEQUENCE 362 AA; 40613 MW; 42DB45A7F72F3B89 CRC64;

Query Match	5.3%;	Score 94;	DB 5;	Length 362;
Best Local Similarity	21.7%;	Pred. No. 6.4;		
Matches 48;	Conservative 27;	Mismatches 88;	Indels 58;	Gaps 9

```

73 1MHAIGRSLTTRKNNTMWEYELF-----QJNCFFPHLR-----PEMD 112
48 VEDDSQSTTRKKNPFSEFEYQGFPEVDNDQVTKRLNSTPIPHRYITQDFLOPIBDM 107
113 AEFQCNQAGACEFEGIDVHKENGELTVQATISGNNENQAKVKNQDNETGLEYETWNV
108 GPFVSV-----TLVFAIGRGN-----LAQIENDGAKGTGSDFRM 145
173 KASEKAE-----TWEDSYDCS--KVLATFNKLAFFGAEFKNIEFTNYRILPYSGPT 225
146 AKIPENYTEISIFRQPPKRNCKLIDFSMKKFOKINMLCESFKKSD-----FCHQNSPS 200
226 YLGNETSVPFGTKTKLGLAIKR---FYRPFKEHLPRKEFL 263
201 FC-MLSHGFLTNGVLTKGTSBQDQNNFYTSFTSASTLILFL 240

```

5330

SUBMIT 14		
025330	PRELIMINARY;	PRT; 1238 AA.
025330;		
01-JAN-1998	(TRIMBLrel..05, Created)	
01-JAN-1998	(TRIMBLrel..05, Last sequence update)	
01-JUN-2003	(TRIMBLrel..24, Last annotation update)	
	Hypothetical protein HP0609.	
	HP0609.	
	Helicobacter pylori (Campylobacter pylori).	
	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;	
	Helicobacteraceae; Helicobacter.	
	NCBI_TaxID=210;	

SEQUENCE FROM N.A.  
STRAIN=26695 / ATCC 700392;  
MEDLINE=97394467; PubMed=9252185;  
Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
Nelson K., Richardson J., Zhou L., Kirschner E.F., Peterson S.,  
Loftus B., Quackenbush D., Dodson R., Khalak H.G., Glodek A.,  
McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Watney L., Wallin E.,  
Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
Venter J.C.;

"The complete genome sequence of the gastric pathogen *Helicobacter pylori*."

Nature 388:539-547 (1997).

EMBL: AEO00575; AAO07677.1, -

PIR: A64596; A64596.

TIGR: Hp0609; -.

DR GO: 0009279; C:external outer membrane (sensu Gram-negativ. . .; IEA  
DR GO: 0015070; F:toxin activity; IEA  
DR InterPro: IPR04311; P:cut\_vacu\_cytox.  
DR Pfam: PF03077; VacA2; 3.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1238 AA; 135062 MW; 66DF754EB1BB173 CRC64;

Query Match 5.3%; Score 94; DB 16; Length 1238;  
 Best Local Similarity 23.6%; Pred. No. 32;  
 Matches 62; Conservative 28; Mismatches 77; Indels 96; Gaps 15;

```

QY      80  FRLSLTGKNTMTEMYELFQLGNCFEPHRLREMDAPFCN-----QCAACFEGGIDDPHWK 134
      1  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      513  FLYRYGGKNSALVF-----NATTP-----WANGSIPKNSVTREFGYEGVNWG 555

```

153 ENGILIVQVAL-----ISGNNEQQMARWVKQDNEIGIYEIMNVKASPERGAEFTWFDSDC 198  
 :   :         |||                 ||  
 Db 556 KTGITGTFTADRVYITGNM-----TC-----NGAQTGGGA--T 588

Db 589 LNFVGALEINIA--GATFKNLKTSQNSYNTFMALGD-----SSGSAKINVSQ 634

Db 635 SDEYDWTGGGYDFTGN-----GVFDSVNFNKAYYKFGTENSYNFCKTNFLAGN 683

Db  
684 FKQGKTTIEKSVLSDASYTFDG 706

RESULT 15	
Q8WHW9	
ID Q8WHW9	PRELIMINARY;
AC Q8WHW9.	PRT; 2313 AA.

DE	Hypothetical protein.
DT	01-OCT-2003 (TREMBLrel. 25, last annotation update)
DT	01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT	01-MAR-2002 (TREMBLrel. 20, Created)

05 *Ptilotum nudum* (Whisk fern).  
06 Chloroplast.  
0C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CO	NCBI_TaxID=3240;
RN	[1]
RP	SEQUENCE FROM N.A.

RA Wakasugi T., Nishikawa A., Yamada K., Sugiyama M.;  
RT "Complete nucleotide sequence of the chloroplast genome from a fern,  
RT *Ptilotum nudum*.";  
Submitted (7 May 2003) to GenBank as F045921.1

DR EMBL; AP004638; BAB84296.1; -.  
DR GO; GO:0009507; C:chloroplast; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0005351; F:sugar porter acti-  
ttr: ttr

DR G6; G6:00059401; P:phosphoenolpyruvate-dependent sugar phospho. . . ; IEA.  
DR InterPro; IPR003959; AAA\_ATPase\_centrl.  
DR InterPro; IPR008543; DUF825.  
DR InterPro; IPR002114; HPr\_Serp\_S  
DR InterPro; IPR008543; DUF825.

DR Pfam; PF05695; DUF825; 1.  
DR PROSITE; PS00589; PTS\_HPR\_SER; 1.  
KW Hypothetical protein; Chloroplast.

Query Match	5.3%;	Score 94;	DB 8;	Length 2313;
Best Local Similarity	21.7%;	Pred. No. 72;		

79 GFRSTLTGKNYTMWYELFOLGNCFTPHLRPEMDAPFWCNGAACEFECIDVH----WK 134

2b 1059 GRSFIRKRGFCLEGTNPFQIKSHLFNKI-PNKEYPH-----SILTQTLHRYLIK 1.11  
2y 135 ENGLIVQVATISG---NMENOMAKWQDNFTGIYYETWVKASPEKGAETW---PDSYD 188  
2b 1112 EFGSSIQLKSLSTEQVNLFDLQERFL--NSSIRKQLVNIQVS-----DYWQPLDSDP 1.63  
2y 189 CSKFLV---RTENKLAERGALEFNKI-ETNYTRIFLYSGEFTYLGW-----ET 231  
2b 1164 TNDPHLNIISTKQDQLNQGGSGSIIDEKSYHNDYLS--KFGNLEBYDMLFRLKIPEL 1220  
2y 232 SV-FGPTGNKTLGLA-----IKRFYYP--FKPHL---PTKEFLSLQTFDAVIY 275  
2b 1221 SIHFIPODSKISLEKHEFNQDIKNIVYDKNIFRSNLMFNERLNVINTELLRVSTI 1280  
2y 276 HKQFYLFYNEFYWFLPMKF-----PFIKITYEEIPLPIRNKTLSQL 316  
2b 1281 AKKWLFPHEYIIPWFETIEWWYINSAVNTPSEFLNINISQWISNL 1326

Search completed: June 8, 2004, 14:07:07  
Job time : 59.3722 secs